# Conflicting Relations Paradigm: The Effects of A Stimulus Equivalence-Based Approach to Changing Bias

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Robert L. Henery

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Jennifer McComas, Advisor

May, 2018

© Robert L. Henery, 2018

**Acknowledgements**

All acknowledgement and thanks must begin with Dr. Jennifer McComas, whom I have been so very fortunate to have as my advisor for the past five years. Her vision, passion, and leadership have been the most formative part of my experience in this graduate program. I am continually inspired by Dr. McComas's dedication to working with all our incredible students. It would easily take an additional five years to detail all of the ways in which she has supported me in this study and facilitated every aspect of it from creation to completion.

I would like to offer my deepest gratitude to my committee members; Dr. Frank Symons, Dr. Jason Wolff, and Dr. Richard Lee. Thank you for bringing your vision and expertise to this study, and for the time that you have taken from your preposterous schedules to serve on my committee. Your willingness to jump in on a relatively unknown topic and help shape this study is a kindness I intend to pay forward to my future students.

Thank you to my former, and exceptional, middle school students. I hope that to some small degree I was able to enrich your life as much as you were able to enrich my own. Teaching you was one of the great joys (and challenges) of my life, our time together gave me the passion and the context knowledge to pursue a higher degree. I will forever treasure the time we were together.

I would like to acknowledge the work that this study was built on. Particularly the work of Marilia Pinheiro de Carvalho, Julio C. de Rose, Táhcita M. Mizael, João H. de Almeida, and Carolina C. Silveira. I found your studies to be marvelously creative and

inspirational. Thanks to Dr. Lanny Fields, who took the time to council an unknown graduate student a couple of thousand miles away about the intricacies of stimulus equivalence. Your expertise in the field is staggering, and it was a true gift to be able to consult with you. And thank you to Dr. Terry Falcomata and Ashley Bagwell from the University of Texas. I still think about the brisket from visiting you both in Texas. Your contributions were integral to the actual running of this study through the creation of the fantastic computer program we used.

A special thank you to my family for the constant support throughout the entirety of this journey. Long drives to watch grandchildren bailed us out in many a pinch. And a final set of thanks to my co-workers who made even the longest days of teaching and writing enjoyable.

## Dedication

I dedicate my dissertation to my wife. Without her support, in every possible sense of that word, this would not have been possible. No one could ask for a better partner in all things that are life. I am going to allocate only a partial dedication to my two children, Holden and Phoebe. The two of you have made this process unimaginably more difficult. Middle of the night illnesses, colic, a preternatural ability to avoid naps, and lost working days to childcare made this so very challenging. However, you still earn big points for your unconditional love and your ability to bring so much joy and happiness to our house on even the most challenging days. You both have made finishing this paper and degree so difficult, but infinitely more meaningful. I love you both dearly.

## Abstract

The stimulus equivalence paradigm has been used in a small number of studies to examine attitudes and responding towards socially relevant stimuli. Two of these studies attempted to use match-to-sample (MTS) training to establish equivalence relations between a positive stimulus and faces of African descent (de Carvalho & de Rose, 2014; Mizael, de Almeida, Silveira, & de Rose, 2016). In the latter study by Mizael et al. (2016), a redesigned training and testing protocol was employed and all 13 of the participants showed equivalence class formation. Furthermore, the authors offered evidence that procedures based on equivalence and transfer of functions were able to reduce pre-experimental biases demonstrated by participants. The present study was a systemic replication of the training and testing protocol from Mizael et al. (2016) with a novel study population that included 8 East African and 7 Native American elementary age children. The school that these children attended reported frequent and hostile interactions between these groups of students. All 15 children learned relations during matching tasks that would potentially establish emergent relations between outgroup faces and positive stimuli. Fourteen of the 15 children showed equivalence class formation. All 15 children also completed an array of stereotyping and prejudice measures before and after delayed match-to-sample training (DMTS) and testing to detect any generalization of the DMTS training effects beyond the experimental context. Participant performance on those measures suggested little, if any, generalization of training effects. Results are discussed in terms of the utility of the stimulus equivalence paradigm for addressing the challenges related to stereotyping and prejudice.

## Table of Contents

**List of Tables**

## List of Figures

# Chapter 1

## INTRODUCTION

Stimulus equivalence is generally defined as the interchangeability of stimuli within a class (Green & Saunders, 1998; Arntzen, Grondahl, & Eilfsen, 2010). Stimulus equivalence has been described as an effective and powerful approach for altering or changing the behavioral functions of stimuli (Mizael et al., 2016). Plaud et al. (1998) note, "Stimulus equivalence is a behaviorally based approach to exploring and understanding how humans process the relations between these things—these symbolic stimuli—a process traditionally viewed as cognitive in nature but for which behavioral response patterns do emerge" (p.63). Research on stimulus equivalence utilizes stimuli, often meaningless or arbitrary in nature, that through training and testing demonstrate the three key properties of equivalence: reflexivity, symmetry, and transitivity. When these three properties have been confirmed, a class of stimuli is considered to be formed (de Rose, Hidalgo, & Vasconcellos, 2013).

Equivalence classes are established by teaching conditional discriminations to establish relations among individual stimuli to form a group. In typical stimulus equivalence studies, a match-to-sample (MTS) procedure is used in which a subject is first shown a sample stimulus (e.g., A). Different stimuli, termed comparison stimuli (e.g., B and C), are then presented. Explicit reinforcement of "correct" choices is used to train the subject to select stimulus B when presented with stimulus A. The same procedure is used to train the subject to select stimulus C when shown stimulus B. These responses (A-B, B-C) are learned through direct reinforcement contingencies and are

entirely expected (Plaud et al., 1998). Stimulus equivalence refers to the unexpected finding that happens when a subject is shown stimulus A as a sample, and given the choice of comparison stimuli C and D (when controlling for the number of stimulus placements and the location of stimulus placements), the subject will most likely choose stimulus C, despite no prior reinforcement contingencies in effect for that choice (Plaud et al., 1998). Through this equivalence of responding, stimulus combinations act as members of a single class without any direct training to predict this response (Plaud et al., 1995). For example, given three arbitrary stimuli (abstract symbols with no prior meaning to the subject): A, B and C, an equivalence class may be established by training two-term relations between AB and BC. If a class is established through training, emergent relations that were not directly trained are formed between the stimuli: reflexive (which is usually presumed if other relations are established, A-A), symmetrical (B-A, C-B), transitive (A-C), and equivalence (C-A). The group of stimuli can be said to constitute an equivalence class if all of the emergent relations control responding (Sidman, Kirk, & Wilson-Morris, 1985) and the stimuli are fully substitutable for one another (Sidman, 1990, 1994). Research has demonstrated equivalence class formation with verbal and gustatory stimuli (Annett & Leslie, 1995; Hayes, Tilley, & Hayes, 1988), auditory stimuli (Dube, Green, & Serna, 1993), tactile stimuli (Plaud, 1995; Tierney, DeLargy, & Bracken, 1995), and musical stimuli (Hayes, Thompson, & Hayes, 1989). A consistent finding in the research is that equivalence classes are formed by human subjects within and across a wide range of stimuli (Plaud et al., 1998). Children as young as 3 have demonstrated equivalence relations and class formation has been successful with classes

or stimuli as large as 8 members (Plaud, 1995). Stimulus equivalence is considered a robust phenomenon (de Rose, Hidalgo, & Vasconcellos, 2013) and has been studied with an increasing breadth in regards to both research topic and methodological components.

A recent and novel application for stimulus equivalence has been the exploration of the effect of stimulus equivalence on prejudice and bias. De Carvalho and de Rose (2014) and Mizael, de Almeida, Silveira, and de Rose (2016) conducted studies that examined the ability of participants who demonstrated pre-experimental bias related to skin color to create equivalence classes with "positive" symbols and human faces of color. De Carvalho and de Rose found that the majority of participants were unable to demonstrate equivalence class formation. Using a more elaborate stimulus equivalence training regimen, Mizael et al. (2016) found not only that participants were able to demonstrate class formation among these stimuli, but that the stimulus equivalence training and testing and resulting class formation also reduced the observed bias related to skin color.

**Statement of Problem: The Creation of Stimulus Classes with Socially Relevant Stimuli**

The Mizael et al. (2016) study was notable in two regards. First, study participants were able to demonstrate class formation using stimuli that had pre-experimental meaning to the participants. An important experimental aspect of most stimulus equivalence research is the careful selection of stimuli to control for pre-experimental relations among them. In 1996, Zentall and Smeets cautioned,

> "No matter how careful an experimenter might be in selecting stimuli and controlling for possible non-arbitrary relations among them, it may be virtually impossible to avoid competing control by such relations in tests for equivalence involving conditional discriminations. If such extraneous relations come to exert partial or total control over test performance, "spurious" equivalence relations may be found, if they are congruent with expected performance; otherwise, if they are not congruent with the expected performance, then test results may show competing control over selections" (p. 260).

Thus, most studies on the topic are designed with arbitrary stimuli that have no physical commonalities so as to demonstrate the phenomenon of stimulus class formation in the absence of pre-experimental history or relations among the stimuli that might confound interpretation of the findings (see Figure 1 for an example).

**Figure 1. Sample Stimuli.**

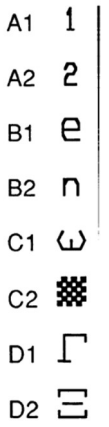



*Figure 1*. Stimulus classes used in Johnson and Sidman (1993). All stimulus letters followed by 1 (e.g., A1, B1, C1, D1) comprised one equivalence class and all stimulus letters followed by 2 comprised a different equivalence class.

Nonetheless, an emerging area within the field of stimulus equivalence research that has been less explored, and intentionally breeches the preceding caution, is the use of "socially relevant stimuli" or stimuli that carry a strong pre-experimental history for study subjects. The design of these studies, termed by Mizael et al. (2016) as the Conflicting Relations Paradigm, is notable not for the demonstrated equivalence of these stimuli but rather for the inability of these stimuli to demonstrate the expected equivalence given the use of established study methods. A more substantial overview of this body of work will be provided in the following chapter, but a commonality among these studies is that participants consistently fail to demonstrate class formation among sets of stimuli that contain one or more stimuli purposely selected because they have prior meaning to study participants. Although the body of research in this area is small, researchers routinely find that expected equivalence relations fail to emerge following training procedures using these socially relevant stimuli. The results of the Mizael et al. study break with this pattern as all 13 of the participants in the study were able to demonstrate class formation, a result the authors attribute in large part to a more robust training protocol. Further research needs to be conducted to examine the effectiveness of this "optimized" training protocol to produce consistent stimulus class formation using stimuli that have a pre-experimental history or are socially relevant to study participants.

The Mizael et al. (2016) study is also notable in the finding that successful stimulus class formation produced a change in the racial bias exhibited by study participants. The authors note the temptation to speculate that stimulus equivalence procedures could prove a valuable tool to fight racism while acknowledging that the

experimental effects of the procedures are likely to be short lived and unlikely to generalize outside of a laboratory setting. The reduction of racial bias seen in the Mizael et al. study represents a significant claim to the effectiveness of stimulus equivalence procedures, a claim of such potential importance that further investigation into how bias was measured and reduced as well as the reproducibility of the finding to additional populations and contexts should be conducted.

**Purpose of the Study and Research Questions**

This study aimed to build on the Mizael et al. (2016) through: (1) examining bias and stereotyping between two culturally distinct groups of students in one elementary school building within a large urban school district in the Midwest, (2) utilizing the "optimized training parameters" found in the Mizael study with two novel groups of students to determine whether equivalence classes could be formed, and (3) conducting pre and post-test measures to examine the generalization of stimulus class formation on participant attitudes towards students from the other cultural group. This study was designed around two specific groups of students, one group consisting of Native American students (primarily from the Anishinabe and Dakota/Lakota tribes) and one group consisting of East African students (primarily Somali). One school building in the district houses two school programs. One program serves primarily East African students with the majority of students identifying as Somali, with the most recent district data showing that 86.6% of the students in that program are identified by parents as African American compared to 38.1% of students across the district (School District data, 2017). The demographic label of African American in the school district encompasses recent

African immigrants, making further distinctions about the size of the East African population in a particular school difficult. The higher percentage of African American students at this school program relative to the rest of the school district as well as an English Language Learner program centered around Somali, Amharic, and Oromo speakers are indicators of the large population of first and second generation East African students in the program. The other school program in the building is a program whose student body is composed mainly of Native American students. The 2017 district demographic report listed 73.3% of the total student body at the school as identified by parents as Native American. The program has been designed as a program to engage urban indigenous students and has a much higher percentage of Native American students compared to the wider school district (73.3% vs 3.5%; School District data, 2017). During the school day the students are separated in the building in their respective school programs. During school transitions such as bus time, recess, before and after school in shared hallways, as well as in the after-school program, the two student populations are able to interact. These interactions have been identified by school and district personnel as a source of conflict between the two groups of students with numerous fights, instances of racial harassment, and frequent disciplinary actions resulting from interactions during this time. Students from both groups report being frequently subjected to bullying, harassment and threats from the other group. School staff have attempted to address what they believe are unhealthy and hostile racial attitudes among the student populations towards different cultural groups of students through social skills training and group work during advisory periods. School staff have reported that these efforts

have shown little to no change in student behavior, nor a lessening of student conflict during the times the groups interact.

This study pre-tested students from both groups to determine their pre-experimental attitudes towards students from the other school program. Pre-testing used the same Self-Assessment Manikin (SAM) measure used by both Mizael et al. (2016) and de Carvalho and de Rose (2014) to document bias. Five additional measures used in contemporary research in the area of stereotyping and prejudice were also administered prior to stimulus equivalence training to examine participant attitudes towards students from the opposite school program. Next, participants completed delayed match to sample training that was designed to establish relations between faces that resemble students from the other cultural group and positive symbols. After participants reached criterion for mastery during training, emergent relations tests were conducted. Following stimulus training and testing, study participants were given the same stereotyping and prejudice measures they completed prior to training to assess any changes in their attitudes towards the cultural group from the other school program. One limitation of the Mizael et al. study, acknowledged by the authors, was the absence of demonstrated generalization of the effects of the stimulus classes formed. The pre and post-test measures in this study attempted to further address and examine the issue of generalizability of stimulus equivalence training by directly measuring students' stereotypes, attitudes, behavioral intentions, and their desire for future contact with the opposite cultural group before and following the training.

The following research questions guided this stimulus equivalence study:

1. Can the optimized training parameters for stimulus equivalence training used by Mizael et al. (2016) be extended and results be reproduced with East African and Native American students in a semi-segregated school environment?
2. What is the extent of generalization of stimulus class formation as indexed by measures of stereotypes and prejudice?

Variables

IV: Delayed match to sample training

DV 1: Results of equivalence tests (demonstration of symmetry, transitivity, and equivalence) and the emergence of stimulus classes following delayed match to sample training (pre- and post-test performance). Although this can be examined as a process in smaller sample sizes (how many trials to reach mastery), for the purposes of the analysis in this study results were a binary yes/no to the participant's demonstration of equivalence class formation. The criterion for demonstrating equivalence class formation was based on the Mizael et al. (2016) study and set at 15/16 or 16/16 correct responses in either the first or second attempt at both the AC and CA tests.

DV2: Participant responses to the five instruments used to measure stereotypes and bias, measured both before and after stimulus equivalence training (Trait Ratings Measure, Intergroup Attitude Measure, Ingroup and Outgroup Behavioral Intentions Measure, Desire for Future Contact Measure, and the SAM).

# Chapter 2

## REVIEW OF THE LITERATURE

The use of socially relevant stimuli in stimulus equivalence research has been a small and relatively recent line of inquiry. It was first explored by Watt, Keenan, Barnes, and Cairns in 1991. The researchers sought to examine the effect of pre-experimental "social training" for subjects by attempting to create equivalence classes with socially significant stimuli. Watt et al. (1991) hypothesized, "…aspects of the relations between the stimuli and the responses have already been …" [learned] "in the social context. Consequently, the researcher focuses upon the end products of the processes of social categorization, rather than on the processes themselves. A behavioral analysis would consider the identification of these processes to be the main object of inquiry….Given that behavior in these settings stems from manipulable antecedent histories, the employment of socially pertinent stimuli might uncover the principles which determine characteristic responding to particular categories of social stimuli" (pg. 34).

The Watt et al. (1991) study took place in Northern Ireland and was designed to examine pre-experimental biases through the specific social context of the "the Troubles," otherwise known as the Northern Ireland conflict. The Troubles lasted from 1969 until 1997 and was a conflict between nationalists (Irish and/or Catholics) and unionists (English and/or Protestants) (Holland, 1999). The Watt et al. study was composed of three subject groups; Northern Irish Catholic subjects, Northern Irish Protestant subjects, and English Protestant subjects. Watt et al. hypothesized that the long

and bloody conflict between these groups had established strong social meaning to the religious stimuli used in the study, and that this deep social meaning might inhibit attempts to use reinforcement to train relations between stimuli that were socially incompatible based on the research participants' social training.

The three groups of university students were trained to match Catholic names (A) to nonsense syllables (B). The subjects were then trained to match the same nonsense syllables to Protestant symbols (C). Given the established body of work on stimulus equivalence and the training procedures used in the study, participants should have demonstrated the emergence of AC and CA relations between Catholic names and Protestant symbols. Watt et al. (1991) found that the English subjects in the study were able to demonstrate the intended equivalence between the stimuli but very few of the Northern Irish participants, either Protestant or Catholic, could do the same. This result suggested that the pre-experimental history between the Northern Irish participants and these religious stimuli interfered with the expected emergence of equivalence between these socially relevant stimuli.

The Watt et al. (1991) study spawned a tangential line of stimulus equivalence research that departed from traditional equivalence research in two important ways. The first was the use of meaningful stimuli that held strong social connotations for research subjects rather than a series of abstract and relatively meaningless symbols. The second was the inability of subjects to create equivalence class formation using these stimuli despite using study methodology that had been repeatedly shown to promote class formation. In a 1995 study comparing the formation of stimulus classes that contained

either fear-relevant stimuli (snake-related stimuli) or fear-irrelevant stimuli (flower related stimuli), Plaud argued,

> "Perhaps one reason why the relationship between stimulus class and stimulus equivalence has not been previously studied has to do with the potential theoretical and procedural complexities that might be involved in studying this hypothetical relationship. Experimental procedures that use nonsense syllables or Greek characters to study equivalence formation do not need grand theoretical models in order to explain behavioral findings, because studies using arbitrary stimuli do not vary on any underlying stimulus dimension" (p. 210).

Examples of the conflicting relations paradigm include the effects of sexually explicit words on the formation of equivalence classes (Plaud et al., 1995), the use of threat words with clinically anxious subjects (Leslie & Tierney, 1993), the use of "terrorist stimuli" following the September 11th attack (Dixon, Rehfeldt, Zlomke, & Robinson, 2006), and gender-role stereotyping (Moxon & Keenan, 1993). One notable study that attempted to create stimulus equivalence came from Haydu, Carmargo, & Bayer (2015). The authors recruited zealous soccer fans in Brazil (criteria for study participants included having at least one tattoo of their favorite team) and attempted to form an equivalence class with symbols of their hated soccer rivals and words like "good" or "champion" as well as an equivalence class containing symbols for their favorite team and words like "bad." The authors found that subjects were not successful in forming equivalence classes. In all of the above examples, study subjects were unable to establish equivalence classes with stimuli that had a meaningful and oppositional pre-experimental social context.

**Training Schedules**

Given that the use of socially relevant stimuli in previous stimulus equivalence research has consistently been shown to impede stimulus class formation, the method of training these relations carries an increased importance. The design of the class size and number of stimuli used in the research can impact the ultimate effectiveness of the training procedures (Plaud, 1995). The greater the number of stimuli and the size of the attempted classes the more difficult class formation will be. For example, attempting to form a class that consists of three stimuli A, B, and C should prove easier and require less training to form than attempting to create a class containing eight stimuli. Using the example of a class of three stimuli, a participant would earn reinforcement when they match stimulus B to stimulus A when stimulus A is the presentation or sample stimulus. Next, they would earn reinforcement when they match stimulus C to stimulus B when stimulus B is the sample stimulus. In this example, the training protocol may consist of one block of the AB training and one block of BC training. Each block would be composed of a set number of trials, generally around 10 to 15 trials that each provide the opportunity for the participant to earn reinforcement for matching the sample stimulus to the correct response stimulus. A participant would be required to meet an established criterion of accurate responding before moving to the subsequent block. Tests for symmetry (B=A, C=B), transitivity (A=C), and equivalence (C=A) would be conducted following the training protocol to verify the successful establishment of the three member stimulus class.

The sequence and dosage of the match to sample (MTS) training can vary greatly from one study to the next (Mizael et al., 2016). In this context, dosage can refer to the number of trials in a training block, or the number of training blocks in the protocol. Examining three of the more recent stimulus equivalence studies will illustrate the variation in training protocols that have been used to attempt class formation with socially relevant stimuli. The first study was by Haydu, Camargo, and Bayer (2015) who attempted to create equivalence classes using soccer club emblems and words such as "good," "bad," and "champion" with Brazilian soccer fans. The second study was conducted by de Carvalho and de Rose (2014) and applied stimulus equivalence to the topic of racial bias. The third and final study to be detailed was done by Mizael et al. (2016) and was a replication of the de Carvalho and de Rose study with a modified training protocol. All three of these studies used differing design components such as the order that training blocks are presented in, the total number of trials in each training block, and at what point in the protocol researchers conducted testing blocks.

In the soccer club stimulus equivalence study, Haydu, Camargo, and Bayer (2015) attempted to create three classes of stimuli with an A, B, and C stimulus in each class. Table 1 shows the training and testing protocol used in their attempt to create the stimulus classes. In each block, participants were required to achieve a criteria of 90% correct responding to advance to the next block, failure to reach criterion in any block required the participant to repeat the block until criterion was met. The training AB 100% and training BC 100% blocks provided participants with feedback after every response.

The term feedback indicates both visual confirmation of a correct or incorrect response as well as a visual tally of a point following a correct response.

Table 1

*Training and Testing Protocol for Haydu, Camargo, and Bayer (2015).*

| Block and Percentage of Trials Reinforced (no trials reinforced during tests) | Conditional Relations | Number of Trials |
|---|---|---|
| Pretraining | X1Y1, X2Y2, X3Y3 | 9 |
| Training AB 100% | A1B1, A2B2, A3B3 | 18 |
| Training BC 100% | B1C1, B2C2, B3C3 | 18 |
| Training AB 50% | A1B1, A2B2, A3B3 | 18 |
| Training BC 50% | B1C1, B2C2, B3C3 | 18 |
| Test Symmetry BA CB | B1A1,B2A2,B3A3,C1B1, C2B2, C3B3 | 18 |
| Test Transitivity AC | A1C1M A2C2M A3C3 | 9 |
| Test Transitivity AC +1 | A1C1, A2C2, A3C3, A4C4 | 16 |
| Test Equivalence CA | C1A1, C2A2, C3A3 | 9 |
| Test Equivalence CA +1 | C1A1, C2A2, C3A3, C4A4 | 16 |

Points accumulated during the training blocks could be used by participates to earn prizes. In providing feedback in 100% of the trials in the initial AB and BC training blocks, correct responding was put on a continuous schedule of reinforcement (CRF). The subsequent AB and BC training blocks reduced the feedback provided to participants to 50% of the trials to promote the maintenance of correct responding to later testing blocks by thinning the schedule of reinforcement.

The final five blocks tested the emergence of relations through demonstration of symmetry, transitivity, and equivalence. The stimuli A4 and C4 were used in the AC +1 and CA + 1 testing blocks to examine responding with the addition of novel stimuli. Stimulus A4 was the emblem of a Brazilian soccer club that was selected to have "neutral emotional functions". Stimulus C4 was the word "champion". During the testing blocks no feedback was provided and there was no repetition of blocks to meet a set criterion. Only one of the 28 participants in this study was able to demonstrate formation of the intended equivalence classes.

De Carvalho & de Rose (2014) conducted a study designed to investigate racial attitudes among children through the use of the stimulus equivalence paradigm. This study took place in Brazil and had four participants aged 7 to 10 who were identified by the researcher's chosen screening measure to have a negative bias towards black men. One of the participants was described as "brown skinned," one as "black skinned," and two as "white skinned." The goal of the study was to determine whether participants could successfully form two classes of stimuli, one class that included a picture of a thumbs-up (positive stimulus A1), a picture of an abstract shape (B1), and pictures of adult black male faces (face stimuli C1) and a second class composed of a picture of a thumbs-down (negative stimulus A2), a picture of an abstract shape (B2), and another picture of an abstract shape (C2). De Carvalho and de Rose wanted to see if they could reverse the behavioral relations demonstrated by participants in the pretest (matching black faces to a negative stimulus and matching white faces to a positive stimulus) through training designed to generate the emergent relations seen in the stimulus

equivalence paradigm. The methodology in this study incorporated two additional elements not utilized in the Haydu, Camargo, and Bayer (2015) study. The first was the use of delayed match to sample training (DMTS). DMTS incorporates a delay, in this case 1.5 seconds, between the presentation of the initial sample stimulus and the appearance of the comparison stimuli. DMTS is hypothesized to strengthen responding relative to MTS procedures. Other researchers, including Mizael et al. (2016), have used match to sample procedures with socially relevant stimuli and found similar results with and without a delay. Results of these studies suggest that the use of DMTS is at the very least as effective as standard MTS procedures, with many researchers continuing to hypothesize that DMTS increases the strength of responding compared to typical MTS procedures.

The second procedural difference incorporated by de Carvalho and de Rose (2014) was the use of a "bank" of stimuli for their C1 and C3 stimuli. Training was designed to generate responses that would match a positive symbol, a thumbs-up, to a picture of a black face. Rather than using only one face with dark skin as their C1 stimulus, the authors used a rotating bank of five different faces with dark skin. This was done to promote the generalizability of the training by using multiple exemplars rather than having responses tied to the idiosyncratic features of only one face.

Table 2 shows the training and testing schedule for the de Carvalho and de Rose (2014) study. The pre-test and final test presented the participants with either the positive or negative stimulus (A1 or A2) and provided three comparison C stimuli as response options (white face, abstract stimulus, or black face). Compared to the Haydu, Camargo,

and Bayer (2015) schedule there were fewer blocks, fewer trials in each block, and feedback changed abruptly from a CRF schedule to no feedback or reinforcement provided in the relation AB/BC block. In addition, the de Carvalho and de Rose study did not incorporate a training phase which is often used in stimulus equivalence research to teach participants how to respond to the MTS procedures. Perhaps importantly, the final equivalence testing phase asked participants to match one of three stimuli with the initial stimulus. The comparison stimuli offered as choices included not only black faces and the arbitrary shape, but white faces as well. The white face stimulus was not present in any of the equivalence training. Only one of the four children in this study was able to demonstrate class formation in the final test block.

Table 2

*Training and Testing Protocol for de Carvalho and de Rose (2014).*

| Block and Percentage of Trials Reinforced (no trials reinforced during tests) | Relation | Criteria |
|---|---|---|
| Pre-test | C-A | Until responses formed a pattern (3-8 trials). |
| Training AB 100% | A1B1, A2B2 | 100% or 6/6 |
| Training BC 100% | B1C1, B2C2 | 100% or 6/6 |
| Training AB/BC 0% | A1B1, A2B2, B1C1, B2C2 | 100% |
| Test CA with C3 | C1A1, C2A2 | |

Mizael et al. (2016) considered their study as a replication of the de Carvalho and de Rose (2014) study with key changes to the procedures used during training and

testing. Their goal was to determine whether a more robust training and testing protocol could increase the yield of equivalence classes containing relations that contradicted racial biases exhibited in the screening and pretest measures compared to the results from de Carvalho and de Rose. Thirteen participants aged eight to ten were enrolled in the study. Seven of the participants were racially identified as white, six as black. Similar to de Carvalho and de Rose, the aim was to create two classes of stimuli, A1B1C1 and A2B2C2. The first intended stimulus class contained a picture of a thumbs-up (positive stimulus A1), a picture of an abstract shape (B1), and a picture of a black male (face stimulus C1). The second intended stimulus class was composed of a picture of a thumbs down (negative stimulus A2), a picture of an abstract shape (B2), and another picture of an abstract shape (C2). Participants in the original de Carvalho and de Rose study were unable to demonstrate class formation, a result common in the socially relevant stimuli research. Mizael et al. argued that many of the studies using socially relevant stimuli in which expected relations failed to emerge employed a training protocol which incorporated features that were less than ideal to generate successful stimulus class formation. These features included: “insufficient review of the baseline, starting tests with the most complex relations (transitivity and equivalence), and a lack of symmetry tests” (p. 452). To address these theorized shortcomings, Mizael et al (2016) designed a study with a more comprehensive training protocol, shown in Table 3.

Table 3

*Training and Testing Protocol for Mizael et al. (2016).*

| Phase and Percentage of Trials Reinforced (no trials reinforced during tests) | Relations | Learning Criterion | Trials |
|---|---|---|---|
| Pre-test (AC3) | AC | _____ | 16 |
| Pre-training 100% | XY | 14/15 | 15 |
| Training 1 AB 100% | A1B1/A2B2 | 15/16 | 16 |
| Training 2 AB 100% | A1B1/A2B2 | 15/16 | 16 |
| Test BA Symmetry | B1A1/B2A2 | 15/16 | 16 |
| Training 1 BC 100% | B1C1/B2C2 | 15/16 | 16 |
| Training 2 BC 100% | B1C1/B2C2 | 15/16 | 16 |
| Test CB Symmetry | C1B1/C2B2 | 15/16 | 16 |
| Training Mixed 1 ABBC 100% | A1B1/A2B2/ B1C1/B2C2 | 15/16 | 16 |
| Training Mixed 2 ABBC 100% | A1B1/A2B2/ B1C1/B2C2 | 15/16 | 16 |
| Baseline Review 1 50% | A1B1/A2B2/ B1C1/B2C2 | 15/16 | 16 |
| Baseline Review 2 50% | A1B1/A2B2/ B1C1/B2C2 | 15/16 | 16 |
| Test AC | A1C1/A2C2 | 15/16 | 16 |
| Test CA | C1A1/C2A2 | 15/16 | |
| Post-test (AC3) | A1C1/A2/C2 | _____ | |

Compared with the original de Carvalho and de Rose (2014) study, Mizael et al.'s (2016) training and testing protocol is considerably more extensive both in the number of training blocks as well as the total number of MTS trials required in each block. A pretraining block was implemented at the beginning of the sequence to familiarize participants with MTS training. The AB relation training was repeated with a second full block following criteria being met in the first training block. Immediately following both

AB training blocks, testing for BA symmetry was conducted. The placement of the symmetry testing blocks contrasts from other studies in which testing blocks only followed after all training blocks had occurred. After the criterion was met for BA symmetry, the participants were then given two successive blocks of BC training. Again, symmetry testing for CB relations directly followed the training blocks. ABBC mixed training was conducted in the next two blocks in which 8 trials of AB and 8 trials of BC relations, in a random order, were trained. The baseline review blocks were identical to the mixed training blocks except feedback (reinforcement) was reduced from the CRF schedule used in the training blocks until this point to 50% (Fixed Ratio 2). After participants finished the baseline review blocks, the final three testing blocks began. In both the AC and CA testing blocks no feedback was provided and each test was repeated a second time, regardless of participant performance. The final AC3 post-test was the same as the pre-test, in which one of the A stimuli (thumbs-up or thumbs-down) was presented with the option of matching it to one of the C stimuli (white face, abstract stimulus, black face). This final test was done to replicate the previous work of de Carvalho and de Rose.

Mizael et al. (2016) hypothesized that the previous failure of participants in the de Carvalho and de Rose (2014) study to demonstrate equivalence in testing blocks could have resulted from the use of the additional stimulus in the AC3 testing phase, that the inclusion of white faces as a response option after having not been a part of the training regimen potentially compromised the results. Consequently, Mizael at al. decided to employ multiple tests to check for equivalence relations. The AC and CA tests were used

to determine whether stimulus classes were formed. In the AC test, participants were presented with an initial A stimulus (either the thumbs-up or thumbs-down) and given two response options (either a picture of a black face or a picture of an abstract shape). The CA test presented participants with either a picture of a black face or the picture of the abstract shape and two response options (thumbs-up or thumbs-down). These tests used only stimuli present during the training protocol (A1, A2, C1, C2). Participants were then given a final AC3 test similar to the de Carvalho and de Rose (2014) test. In this test participants were presented with one of the A stimuli (thumbs-up or thumbs-down) and given three response options; a black face (C1), an abstract shape (C2), and a white face (C3). The C3 stimuli, the bank of four white faces, were not part of the training protocol. Mizael et al. found that all 13 of the participants showed successful equivalence class formation in the AC and CA tests, a result which the authors attributed in large part to the more extensive training protocol. However, when the post-test was modified to also include white faces as a response option, only nine of the 13 study participants were able to continue to demonstrate relations among the trained stimuli. Mizael et al. referred to the AC3 test as a "different and more stringent test" of relations (p. 451). It is unclear how the authors conceptualized of this test as "more stringent." It is possible that they viewed this AC3 test as a greater demonstration of relational responding as the response options expanded from two to three stimuli. Had another abstract shape been used to expand the response options, it would seem that this AC3 test could be considered more stringent. However, the nature of the stimuli being used in the study raise questions about the use of this particular test. The majority of the participants in the study were identified

as white. The AC3 test put them in a situation of having to choose to match either a white or black face with a positive stimulus. If one of these participants were to match a face similar in skin color to their own to the positive stimulus, should this be considered an undesirable response? And perhaps more importantly, should this pattern of responding be considered as evidence of racial bias?

**Stimulus Equivalence and Racial Bias**

Both the de Carvalho and de Rose (2014) and the Mizael at al. (2016) studies utilized the conflicting relations paradigm to examine racial bias through stimulus equivalence. De Carvalho & de Rose prefaced their study with the argument that the use of stimulus equivalence to conceptualize and investigate the origins of new behavior has extended the experimental analysis of behavior into new and complex domains of cognition, including attitude formation. They further postulated that the pre-experimental history that impacts equivalence class formation of socially relevant stimuli is largely a product of attitudes formed by either direct experiences in or social practices of his or her community. Importantly, the authors hypothesized that attitudes can be formed or modified through the stimulus equivalence paradigm. The authors pre-tested children by asking them to match faces of black and white men with either positive or negative symbols. Children who repeatedly consistently paired white faces with a positive stimulus and black faces with a negative stimulus were selected to participate. Participants then received MTS training designed to form an equivalence class that included both a positive stimulus and black faces. The study sought to examine whether MTS training with reinforcement could reverse the behavioral relations the children

exhibited during pre-training. The child participants were trained to match a black face to an arbitrary symbol, and then trained to match that arbitrary symbol to a positive symbol. Based on decades of empirical research in the area of stimulus equivalence, the children should have then demonstrated class formation by matching a black face to a positive symbol. Three out four children ultimately were unable to form the equivalence class.

Mizael et al.'s (2016) replication of the de Carvalho and de Rose (2014) study used a similar method to determine pre-experimental racial bias. In the pre-test, participants matched a white face, black face, or abstract symbol to a positive or negative stimulus (picture of a thumbs-up or thumbs-down). Only children who demonstrated a consistent pattern of matching the white faces to the positive stimulus and the black faces to the negative stimulus were considered as exhibiting racial bias and enrolled as participants in the study. The authors also employed a self-assessment manikin (SAM) measure as a way to check for bias. The SAM is an instrument used to measure stimuli along three potential domains: pleasure, arousal, and dominance (Bradley & Lang, 1994; Mizael et al., 2016). Mizael et al. only used the domain of pleasure. Participants were instructed to rate each stimulus (thumbs-up, thumbs-down, abstract shapes, and black and white faces) used in the study by selecting a circle that best represented what they felt when they looked at the stimulus (see figure 2 for the SAM manikins). The authors could then compare the ratings the participants assigned to the white face stimuli and black face stimuli.

The pre-test mean SAM ratings for white faces were significantly higher than those for black faces. Following MTS training and equivalence testing the post-test mean

SAM ratings of the face stimuli showed enough of a convergence that the difference between the white and black faces was no longer statistically significant.

**Figure 2. SAM Manikins**

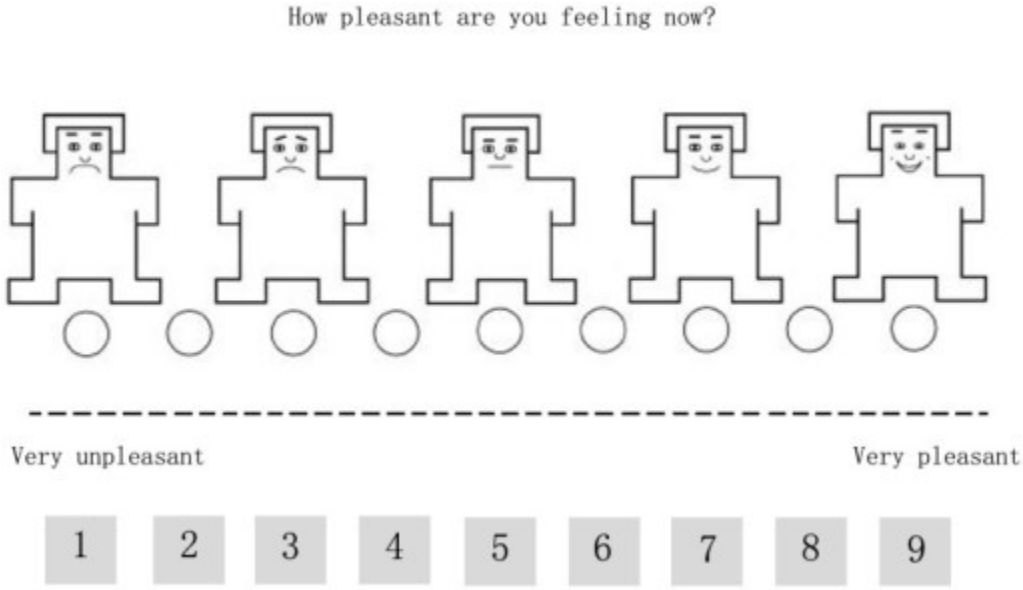



*Figure 2.* Sample self-assessment manikin (SAM). Participants are shown a stimulus (not pictured), and asked to rate their level of pleasure. Participants may select any circle either below or between manikins that corresponds with the pleasure they feel towards the stimulus. The selected circle is then converted to a numerical score ranging from 1-9,

The de Carvalho and de Rose (2014) and Mizael et al. (2016) studies are innovative in both conceptualization and method, as well as intellectually provocative. Specifically, the authors expanded the scope of stimulus equivalence research into the socially important area of racial bias and attitude formation. The importance of addressing racial bias, especially for children in our schools, is immensely important work. The claim that stimulus class formation led to a reduction in racial bias is so

potentially important, and the research poised in such a stage of infancy, that a deeper investigation is needed to address several questions prompted by this research. Although Mizael et al. found differences in pre and post testing using the SAM and the AC3 test, does that in fact provide evidence of a reduction in racial bias? Did the measures employed in these studies capture and measure the construct of racial bias in the first place? Linking the behavioral work of stimulus equivalence to a new domain such as attitude formation necessitates reorganizing the theory guiding the intervention as well as potentially adopting measures more widely used in research that examines racial bias in children. If this work in stimulus equivalence is to be translated to the cognitive domain of attitude formation, and specifically target racial bias, it seems prudent to align both the framework and measures used to the work being done on stereotypes and prejudice in children.

**Stereotypes and Prejudice in Children**

Stereotyping and prejudice are distinct constructs. Racial stereotyping is considered as the "generalized attribution of a set of personal characteristics to members of a racial/ethnic group" (Bigler & Liben, 1993, p. 1507). Racial prejudice was defined by Aboud (1988) as "an organized pre-disposition to respond in an unfavorable manner toward people from an ethnic group because of their ethnic affiliation" (p. 4). Stereotypes and prejudice affect people on multiple levels. When people are subjected to implicit or explicit prejudice they exhibit lower performances on tasks, demonstrate impaired intellectual functioning, and have lower rates of academic engagement (Inzlicht, Tullett, Legault, & Kang, 2011). *Stereotype threat*, apprehension or anxiety felt by group

members who suspect or perceive that their own behaviors and actions will be viewed through a lens of stereotypes, has been shown to produce stress, overeating, obesity, violence, reduced task performance, increased risk-taking behavior, and have long term negative health implications (Steele & Aronson, 1995; Inzlicht et al., 2011). The effects of stereotyping and prejudice, both in explicit actions and perceived threat, can lead to differences in educational attainment, job choice, and career advancement (Inzlicht et al., 2011).

The development of stereotypes and prejudice in children is a topic subject to a number of theoretical models. Social identity theory has arguably been the most influential theory in the field of social group conflict research (Cuhadar & Dayton, 2011). Social identity theory posits that as infants and young children begin to navigate their physical and social world, they use the processes of categorization and evaluation to classify objects and people (Brown et al., 2017; Neisser, 1989). These cognitive processes extend to social development as individuals begin to categorize and evaluate people along salient dimensions. Social ingroups and outgroups are formed as children begin to categorize by grouping people who are “like them” and “unlike them” (Cuhadar & Dayton, 2011). From an early age, children demonstrate ingroup positivity bias, showing stronger positive attitudes towards their ingroup than the outgroup (Cameron, Alvarez, Ruble, & Fuligni, 2001; Brown et al., 2017).

Although developmental researchers agree that categorization and evaluation promote the formation of social ingroup and outgroups, disagreement exists about the importance of these groupings. Children consistently exhibit a positive bias towards their

own ingroup, but favoring one's own ingroup does not necessitate negative outgroup attitudes (Brown et al. 2017). Much of the bias research has focused on these differences using attitude ratings between the ingroup and the outgroup. However, if both groups are viewed positively, with the ingroup viewed more positively, does this meet the threshold of prejudice? Brown et al. (2016) argue that an essential component of prejudice is an expression of negativity toward an individual based on their group membership. Brewer (1999) identified four stages that bridge ingroup identification and intergroup conflict. First, individuals create discrete ingroup and outgroups categories, a process termed *social categorization*. Second, people demonstrate an increased value towards their own ingroup and foster positive relationships with other members of their ingroup (*ingroup positivity*). The element of ingroup positivity does not involve ascribing negative feelings or attributes towards the outgroup. The third element is *competitive intergroup comparisons*, the process of increasing ingroup positivity through social comparisons to the outgroup. The final element, *outgroup hostility*, may develop through conflict between the groups.

Developmental Intergroup Theory (DIT) integrates two theoretical approaches used to explain the development of stereotyping and prejudice in young children; intergroup theory and cognitive developmental theory (Bigler & Liben, 2007). *Intergroup theory* combines aspects of social identity and self-categorization theories to examine the formation of social ingroups and outgroups. *Cognitive-developmental theory* incorporates approaches to cognitive development in early childhood and adolescence. Research in stereotypes and prejudice can be conducted through the manipulation of novel social

groups and assessment of the attitudes resulting from the formation of these groups. Bigler, Jones, and Lobliner (1997) created group identities by assigning children who did not know each other previously to groups that wore different colored t-shirts. In experimental classrooms teachers used the color of the shirts to organize activities, seating, and classroom displays. Control group classrooms ignored the color of the students' shirts. Ingroup biases developed in the experimental classrooms after only four weeks.

Explanations of social stereotyping inherently involve accounting for the process of categorization. Although researchers agree in the existence of this core process, why and how children categorize remains debated. Bigler and Liben (2007) believe that social categorization occurs around perceptually salient features of people such as age, race, gender, and perceived attractiveness. They also identify proportional group size as a dimension that can influence the establishment of stereotypes and prejudice as smaller minority groups appear more distinctive than larger majority groups (Bigler & Liben, 2007). The psychological salience of group distinction is heighted when the basis of group membership is utilized by adults in the context of the environment ("girls sit here, boys line up over there", "3$^{rd}$ graders come with me, 4$^{th}$ graders stay and work"). DIT expects the process of categorization to be influenced by the age or development of the child (classification skills increase through cognitive maturation) as well as environmental experiences such as encounters that further category distinctions. Categorization in turn leads to processes that attach meaning to social groups demonstrated though affect (i.e. prejudice) and beliefs (i.e. stereotypes). Salient physical

differences lead to inferences of other unseen, inherent qualities between groups. My seven year old son recently explained to me that his lunch table had collectively decided that the skin color of each student at the table was determined by how much that child's mother liked to be in the sun, with lighter skinned students having mothers who were in the sun longer. These children had taken note of a salient physical difference between them and attributed it to unseen, inherent qualities of their mothers.

In outlining DIT, Bigler and Liben (2007) identify four factors that are likely to exacerbate the formation of ingroup and outgroup biases in children; perceptual discriminability between groups is exaggerated in a meaningful way, the existence of numeric imbalances in group sizes, explicitly calling attention or labeling groups and using group membership as a basis for actions in the environment, and the creation or perpetuation of conditions like segregation that convey implicit differences between groups. Bigler and Lieben hypothesize that effectively combatting formed stereotypes and prejudice in children might depend on interventions that are able to account for and target beliefs and attitudes towards specific social groups formed in a particular context.

All four of the factors identified by Bigler and Liben (2007) are present in the school setting that this research project was conducted in. The perceptual discriminability of the two groups of students is pronounced. The two groups of students have different skin colors, dress differently (e.g., girls wear hijabs), engage in different cultural practices at school (e.g. prayer, drum circles, cultural ceremonies), and often speak different primary and secondary languages. There is a numeric imbalance of students, with 202 total students in the Native American school and 594 in the other school

(School District data, 2017). Because each program is relatively homogenous, group distinctions are explicitly used as the basis for actions in the school. For example, each time there is a school assembly for the students in the Native American school or a field trip for the other school, the distinction between the groups is highlighted. The final factor, conditions such as segregation, exists for the majority of every school day with students inhabiting the areas assigned to their respective program with little, if any, organized or formal interaction between the groups of students.

**Measures to Study Stereotypes and Prejudice**

De Carvalho and de Rose (2014) and Mizael et al. (2016) sought to examine the effects of stimulus equivalence training on racial attitudes and bias. Both studies used the SAM and their AC3 test as the primary measures to examine racial attitudes and bias. In the case of the SAM, participants were asked to rate stimuli (faces) along the domain of pleasure, with scores ranging from one to nine and each number corresponding to a face on a manikin. Participants who rated stimuli of white faces higher than stimuli using black faces were considered to have a bias against black faces. This bias was considered confirmed though participant responding on a second measure, the AC3 test. The AC3 test was an MTS task where participants were given the positive or negative stimulus (A1 or A2), and provided with three comparison stimuli; a black face (C1), an abstract shape (C2), or a white face (C3). This initial administration of the AC3 test served as a pretest and was administered a second time following the training protocol to determine changes to responses made by participants. Students who repeatedly matched the positive stimulus to white faces and the negative stimulus to black faces were confirmed to have a bias

towards the black faces. Mizael et al. also used an implicit relational assessment procedure (IRAP) as a measure to identify biases and stereotypes. The IRAP requires the participant to select stimuli conditionally similar to the Implicit Association Tests (IAT), in which latency is measured between responses for rapidly changing stimuli (e.g., response time to pair "thin person" with "good" vs. "overweight person" with "good"). Other studies have incorporated multiple measures to converge data to confirm or reject biases (Mizael et al., 2016). Mizael et al. gave the IRAP to study participants only once, following stimulus equivalence training and testing, and found response latency did not reveal either a positive or negative bias towards white or black faces. The authors used this result to corroborate the post-test findings from the SAM that there was no longer a significant difference between participant pleasure ratings for white and black faces.

The measures employed in these studies are quite different from the measures typically used in stereotype and prejudice research. The post-test administration of the SAM with white participants found that pleasure ratings for black faces increased. These ratings followed a number of trials that successfully trained participants to match black faces to a positive stimulus. Assuming that the positive stimulus was seen in a positive regard by participants, the emergence of this relation should produce this change through a transfer of function. This occurs when a function imparted to one stimulus class member is transferred to other stimuli class members without direct training (Mizael et al., 2016). If one stimulus generates positive feelings and affect, when it becomes a member in a class of stimuli, the other stimuli in the class will inherit a degree of similar positive feelings and affect. The pre and post-test differences in the SAM speak to the

effectiveness of the training and successful class formation of the stimuli, but can it really be said to constitute evidence of a change in racial bias? I would posit that researchers in the field of stereotypes and prejudice would argue that the SAM on its own is insufficient to measure racial bias. Mizael et al. noted in their limitations that future research should utilize other and more varied tools to measure racial bias. The measures that are more commonly used in stereotyping and prejudice research instead target differences in attitudes between ingroups and outgroups, generally utilizing multiple measures of attitude. The following measures have been used previously in stereotyping and prejudice research with children.

**Multi-group Ethnic Identity Measure (MEIM).** The MEIM (see Appendix A) is an instrument used to assess affiliation with one's ethnic group (Brown et al., 2014). Ethnic identity is considered to be a multidimensional construct that refers to the quality of a person's affiliation with his or her ethnic group. The development of ethnic identity involves two processes; the exploration of the meaning of one's identity and the sense of commitment and/or belonging to that identity (Marcia, 1980; Ong, Fuller-Rowell, & Phinney, 2010). The MEIM was designed by Phinney (1992) to measure the subjective sense of membership in any group along both the exploring and commitment dimensions. The MEIM is frequently used in research to explore the association between ethnic identity and dependent variables that measure the physical and emotional well-being among racial and ethnic minorities. A recent meta-analysis found 130 studies that have used the MEIM, which is considered to have good reliability with internal consistency (Cronbach's α) ranging from .76 to .91 for the two subscales and .81 to .89 for the overall

scale (Phinney & Ong, 2007; Yoon, 2011). In the present study, participants completed a modified 6 item version of the MEIM. The modified version has been found to retain the same level of reliability as the original version with fewer questions (6 versus 12), some of which have been rewritten to suit younger participants (Homma et al., 2014).

The initial portion of the MEIM is an open ended question meant to elicit a spontaneous statement of one's chosen ethnic label (self-identification). Self-identification is distinct from one's ethnicity and is a precondition for ethnic identity. For this reason, the MEIM begins with the open ended self-identification of one's ethnic label and has been modified to provide examples that are likely to be relevant to the specific population of participants in this study. The open ended self-identification is then followed by 6 questions that use a 4-point scale ranging from strongly agree to strongly disagree for participant responses. The first three questions are designed to measure responses in the exploration domain, the last three questions measure responses in the commitment domain. Scores are derived by summing across items and obtaining the mean. Mean scores approaching 4 indicate high ethnic identity, mean scores approaching 1 indicate low ethnic identity.

**Trait Ratings of Ingroup and Outgroup Members**. Participant attitudes towards members of the ingroup and outgroup will be measured using trait ratings (see Appendix B). Trait ratings measures, in various forms, have been used to study stereotypes and prejudice since 1935 (Katz & Braly). This measure has been used with children as young as 6 years old (Lam & Seaton, 2016) and with a diverse range of populations. Generally, participants are asked to rate how many members of each group

could be characterized by five positive (friendly, good, nice, helpful, and smart) and five negative (dirty, mean, naughty, selfish, and unfriendly) traits taken from the Preschool Racial Attitude Measure II (Williams, Best, Boswell, Mattson, & Graves, 1975). Researchers are then able to examine differences in the ratings attributions between the ingroup and outgroup. Like many of the measures used in stereotyping and prejudice research, trait ratings measures can be adapted to suit characteristics of the populations under study and appear sensitive to change. For example, Vezzali, Stathi, and Giovannini (2012) used this measure to examine stereotyping and prejudice between Italian middle school students and foreign students. Their trait ratings measure was adapted to have only 6 traits; friendly, nice, good, bad-mannered, dirty, and bad. The response scale ranged from 1 (none) to 7 (almost all). Participants in the experimental condition (reading books about students from an immigrant culture) endorsed lower negative intergroup stereotypes than students in the non-intercultural book reading conditions.

**Intergroup Attitude Measure**. Intergroup Attitude Measures (see Appendix C) are used to collect data on generalized attitudes, often in stereotype and prejudice research, as a measure to examine tolerance towards a specific outgroup. Liebkind and McAlister (1999) used this measure to study prejudice of students in Finland towards Russian and African immigrants by having participants respond to items such as, "Foreigners should try and become as much like Finns as possible, even if it would mean that they have to abandon their own language and culture" (p. 773). Response options ranged from 1 (strongly agree) to 5 (strongly disagree). The items included in their measure were taken and adapted from previous stereotyping and prejudice research. To

study the effects of peer modeling on tolerance, McAlister et al. (2000) used an intergroup attitude measure to examine tolerance and prejudice related to five ethnic groups: Whites, Mexican or Mexican Americans, Blacks or African Americans, Asians, and Jews or Jewish people. Examples of items on their measure include, "it would be good to have more ____ here" or "most ___ are similar to my people in values they teach children." The present study used a 5 point response scale similar to the Liebkind and McAlister (1999) study.

**Ingroup and Outgroup Behavioral Intentions Measure.** Behavioral intentions measures (see Appendix D) pose hypothetical contact scenarios and ask participants how they would behave towards ingroup and outgroup members. Hypothetical contact scenarios are modified and adapted for the age of the participants and the specific outgroup that researchers wish to measure attitudes towards. Cameron and Rutland (2006) asked children to imagine they met a student from school with a disability (the type of disability was manipulated). Using a 5 point response scale from "not at all" to "very much so", the children were asked questions such as "How much would you like to play with the student?" or "How much would you like to have them over to play?" Vezzali, Stathi, and Giovannini (2012) employed a similar measure where Italian students were asked to imagine being at park and meeting either an Italian (ingroup) or immigrant (outgroup) student. Items included questions such as, "Are you happy to meet him/her?" or "Would you invite him/her to go and get an ice-cream together?" Differences in ratings between the ingroup and outgroup scenarios can then be calculated

to examine whether there is a greater intention to behave more positively with either the ingroup or outgroup (Vezzali, Stathi, & Giovannini, 2012).

**Desire for Future Contact Measure.** A desire for future contact measure (see Appendix N) can be used to examine participant attitudes towards possible intergroup contact in the future. Tropp and Bianchi (2006) used this measure in studying the effect of explicit references to group membership on feelings towards intergroup interactions. Vezzali, Stathi, and Giovannini (2012) modified the measure to determine participants' attitudes toward the possibility of intergroup contact before and after their reading intervention. Participants were asked if they would like to meet or spend more time with immigrant students and responded on a 7 point scale that ranged from 1 (not at all) to 7 (very much).

# Chapter 3

# METHODS

## Participants and Setting

Fifteen students from a large urban school district in the Midwest United States participated in this study. Participants met the following inclusion criteria; all (1) were enrolled in grades 3-5 in one of two schools that shared a particular building, (2) were enrolled in the after-school program, (3) had their ethnicity identified by their parents to the school district as being either Native American or African American. Given that the school district only uses the ethnic identifier of African American, student self-identification on the MEIM was used to further distinguish East African (Somali, Ethiopian, Oromo, Eritrean) participants for the study. There were no instances of a

participant being identified as African American to the school district and not self-identifying on the MEIM as East African. Table 4 shows the grade, gender, and ethnicity of each participant according to district data for Native American participants and according to both district data and MEIM self-identification for East African participants. Seven participants were identified by parents to the district as Native American, eight were identified to the district and self-identified on the MEIM as East African. Nine participants were identified to the district as female, and six as male.

Table 4

*Study Participants and Demographic Info*

| Participant Number | Gender | Grade | Ethnicity |
|---|---|---|---|
| 1 | F | 5 | Somali |
| 3 | F | 5 | Somali |
| 4 | F | 3 | Somali |
| 7 | F | 5 | Somali |
| 8 | F | 5 | Somali |
| 14 | F | 4 | Somali |
| 15 | F | 5 | Somali |
| 23 | M | 5 | Ethiopian |
| 2 | F | 3 | Native American |
| 5 | M | 3 | Native American |
| 10 | M | 4 | Native American |
| 12 | M | 4 | Native American |
| 13 | M | 4 | Native American |
| 16 | M | 5 | Native American |
| 19 | F | 3 | Native American |

Pre-tests, DMTS training, and post-tests were all conducted during the after-school program at the building so that students did not miss instructional time during the school day. The majority of data collection occurred in one of the school's computer labs, with DMTS training and testing completed on school computers using an DMTS program created in Python and run through the PyCharm application.

**Stimuli**

A selection process was used to identify the stimuli used in the stimulus equivalence procedures. Both Mizael et al. (2016) and de Carvalho & de Rose (2014) used an image of a hand making a thumbs-up as the "positive" stimuli and the image of a hand with a thumb pointed down as the "negative" stimuli. Because the study in this paper worked with novel student populations, preliminary stimulus selection pilots were used to validate the key properties of the stimuli. The stimulus selection pilots consisted of 19 students who had been identified by their parents as either Native American or African American and further self-identified as East African (10 Somali students and 9 Native American students) to the school district, were not participants in the study but were of a similar age and grade to the study participants (grades 3-6). All procedures used with students who participated in these stimulus selection pilots were conducted in private school spaces in school building with each student individually. Results were aggregated for each group of students. The stimulus selection pilots were asked to complete three tasks to identify the stimuli used in the study. First, students from both groups were presented with ten different positive images (e.g. smiling face, thumbs-up, happy-face emoji, the Facebook "like" symbol, a happy animal, a rainbow, etc.) with the direction "please choose a picture that you would pair or match with a person that you like, that you would want to spend time with, that you think of as a friend." After the student made their first selection, the experimenter asked the student to make a second, third, and fourth choice with the direction repeated each time. The first selection made by a student received a score of 4, the second choice a score of 3, the third choice a score of

2, and the fourth choice a score of 1. Total scores from each student stimulus selection pilot were summed to determine which positive stimuli were most frequently and definitively selected by the group of Native American and East African students. Scores from both student stimulus selection pilots supported the use of the same positive stimulus (a smiling yellow emoji with two thumbs up) to be used as the A1 stimulus, shown in figure 3.

Next, the selection of the A2 negative stimulus was determined with a similar selection process to the positive A1 stimulus by the stimulus selection pilots. Presented with 10 images (e.g. face making a frown, thumbs-down, mad emoji, the Facebook "dislike" symbol, etc), the students were instructed to "choose a picture that you think is best to pair with a person that you do not like, that you would not want to spend time with, that you think of as unfriendly." Again both stimulus selection pilots selected the same negative stimulus (a frowning yellow emoji with a thumbs-down) shown in figure 3.

The final task for the stimulus selection pilots was to select the faces used as stimuli to represent the student ingroups and outgroups. The stimulus selection pilot of Somali students was given 27 pictures of student faces and asked to "choose four faces that you believe most people would think of as being East African or Somali, that you think most of your classmates would think are East African or Somali, that you think look most like an East African or Somali student." The faces presented to the stimulus selectionpilot were of children that appeared to be in the age range of 7-14 years old, with a range of skin tones, dress, and genders represented in the photos. Each selection of a

face gave that stimulus one point. Responses were summed to determine the faces used as the C1 stimuli. Similar to the Mizael et al. (2016) study, C1 and C3 were each comprised of a "bank" of four faces to prevent stimulus control to the idiosyncratic features of any one face. The use of multiple stimuli for C1 and C3 is supported in research as a way to strengthen the potential generalization of the training by training multiple exemplars of a perceptual class of stimuli (Fields et al., 2002). The four East African faces were comprised of the two highest scoring male and female faces, to avoid the possibility of stimulus control to the perceived gender of the faces used as stimuli. The same selection process was applied to the Native American stimulus selection pilot to select the faces that most resembled Native American students and were included as the stimuli for C3.

Figure 3 depicts all the stimuli used. The goal of the DMTS training was to establish the formation of two classes of stimuli. For East African students, the two intended classes were A1B1C3 (positive stimulus, abstract shape, and Native American face) and A2B2C2 (negative stimulus, abstract shape, abstract shape). For Native American students the two intended classes were A1B1C1 (positive stimulus, abstract shape, East African face) and A2B2C2 (negative stimulus, abstract shape, abstract shape). B1, B2, and C2 were arbitrary stimuli of geometric shapes intended to have no inherent pre-experimental meaning to participants. The inclusion of the arbitrary geometric shape C2 was to avoid creating a class of stimuli that included the negative stimulus and any of the face stimuli. To allow for time to finalize the computer program that used the selected stimuli, the procedure to select stimuli was conducted approximately two months before DMTS training and testing began

**Figure 3. Stimuli Used in the Training and Testing Protocol**

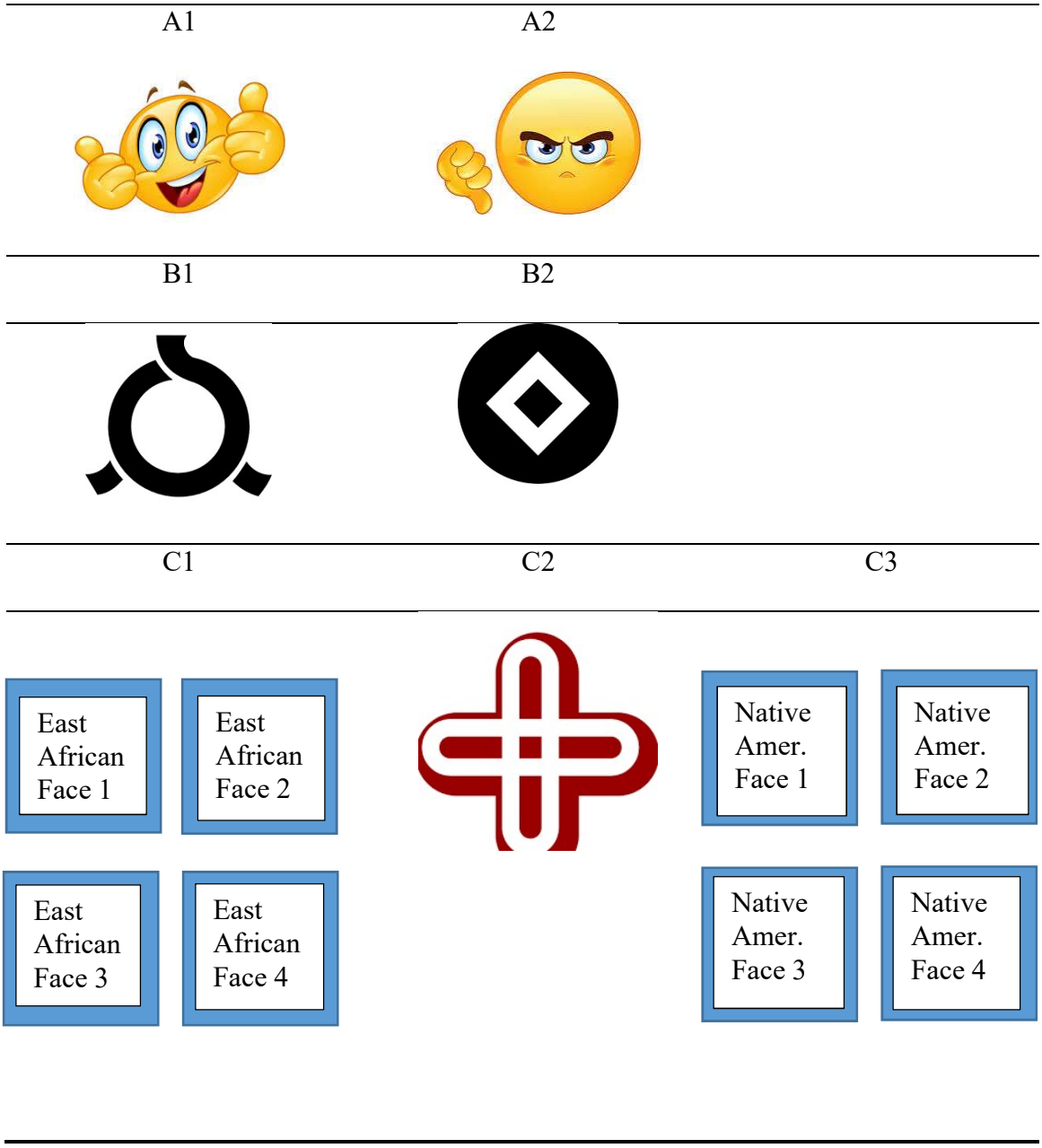



*Figure 3.* Stimuli selected by stimulus selection pilots for use in DMTS. Training procedures were designed to establish relations between; A1 and B1, B1 and outgroup faces, A2 and B2, and B2 and C2). Pictures used as ingroup and outgroup faces have been omitted from this paper to ensure the protection of privacy.

## Assessment Instruments

**Multi-group Ethnic Identity Measure (MEIM).** The initial pre-intervention measure participants completed was the 6 item adapted version of the MEIM. The MEIM began with an open ended question that asked participants to self-identify their ethnic group. Participants were then given 3 statements designed to measure exploration of their ethnic identity and 3 statements designed to measure commitment to their ethnic identity. For each statement, participants responded along a four point scale ranging from strongly agree (4) to strongly disagree (1). Scores for the two subscales and the overall scale were calculated by averaging the item values. See Appendix A for the MEIM measure and Appendix I for screenshots of the MEIM as it appeared to participants.

**Trait Ratings of Ingroup and Outgroup Members**. Participant attitudes towards members of the ingroup and outgroup were assessed using a trait rating measure. Participants were asked to rate how many members of each group could be characterized by five positive (friendly, good, nice, helpful, and smart) and four negative (mean, naughty, selfish, and unfriendly) traits taken from the Preschool Racial Attitude Measure II (Williams, Best, Boswell, Mattson, & Graves, 1975). Responses for each group were scored on a 4 point scale, with positive traits reverse scored. Scores ranged from all (4) to none (1). Writing the questions to target ingroups and outgroups presented a challenge. One option included direct ethnic phrasing of each question. For example, “How many of the Somali students are mean” or “How many of the East African students are friendly?” Prior to conducting this measure, a variety of forms for these questions were posed to students who were not study participants. Some students expressed confusion toward this

use of direct ethnic phrasing. One East African student asked what Native American meant. However when she was asked which program she was enrolled in she pulled at the hijab she was wearing and asked, “does it look like I go to Wiconi School (the Native American program)?” This statement implied that this particular student was aware of salient differences between the groups, and that groups were orientated around program affiliation. Based on this information gathering, it was decided that the clearest way to phrase the questions was to use program affiliation as the way to orient the questions to the ingroup and outgroup. Given that 73.3% of the students in the Native American school (written here as *Wiconi school* to preserve confidentiality and privacy) were Native American and 86.7% of the students in the other school (written here as *Canug school* to preserve confidentiality and privacy) were African American (overwhelmingly East African students), the questions became “How many of the *Wiconi school* students are unfriendly” or “How many of the *Canug school* students are selfish?” Responses for positive traits were reverse scored so that higher scores reflected more negative stereotypes See Appendix B for the trait ratings measure and Appendix J for the trait ratings measure as it appeared to participants.

**Intergroup Attitude Measure.** To measure participant attitudes toward the outgroup, 11 items were adapted from the intolerance scale items used by Liebkind and McAlister (1999). Some items on the intolerance measure used by Liebkind and McAlister, such as, “many foreigners come to Finland just to exploit our social benefits,” were not retained and adapted as they were not applicable to students in this social context. The 11 items that were retained were adapted to fit the context of this particular

school setting. Similar to the trait ratings measure, reference to the outgroup was done by referring to students in the other school program such as, “Wiconi/Canug students cause a lot of problems like fights and bullying in this building.” For each of the items, the word “outgroup” was replaced in the item with either “Wiconi students” or “Canug students” based on the school program of the subject being questioned. Responses were scored on a 6 point scale ranging from very untrue of what I believe (1) to very true of what I believe (6). Examples of items included “I would like it if there were more Wiconi students in the building?” and “I do not want Wiconi students in my classroom?” (reverse-scored). Responses were coded so that higher scores reflect more positive intergroup attitudes. See Appendix C for the intergroup attitude measure and Appendix K for the intergroup attitude measure as it appeared to participants.

**Ingroup and Outgroup Behavioral Intentions Measure**. To assess how participants would behave toward ingroup and outgroup members in a hypothetical contact scenario, four items were adapted from Cameron and colleagues (Cameron & Rutland, 2006; Cameron et al., 2006, 2007) as well as Vezzali, Stathi, & Giovannini (2012). Participants were asked to imagine a new student whom they do not know arriving in their classroom, either an ingroup or an outgroup child. Each participant responded to all four items for an ingroup scenario and then completed the same four items for an outgroup scenario. Participants were then asked the following questions: “How happy are you to meet him/her,” “How happy would you be to play with them,” “How happy would you be to sit next to them at the lunch table,” and “How happy would you be to work on a class assignment or do schoolwork with this person?” For each

question an image was presented of either an East African or Native American child or adolescent. This was done to increase the salience of item towards either the ingroup or outgroup. Images were chosen based on scores from the stimulus selection pilots that were used to select the stimuli for the stimulus equivalence training and testing in the study. The images used for this measure were the highest scoring images from each stimuli pilot that were not used as stimuli in the training and testing. Male participants were presented with images of male students and female participants with images of female students. This was done to control for the influence of gender in responding (a third grade female student may not feel comfortable sitting at lunch with a male student from any ethnic group). Participants responded to the questions using one of the following four choices: very happy (4), happy (3), unhappy (2), or very unhappy (1). See Appendix D for the full ingroup and outgroup behavioral intentions measure and Appendix L and M as the ingroup and outgroup behavioral intentions measure appeared to participants.

**Desire for Future Contact Measure.** Three items, adapted by Tropp and Bianchi (2007), were used to assess the desire for future contact with outgroup members: "Would you like to meet more outgroup members," "Would you like to spend more time with outgroup members," and "Would you like to get to know outgroup members?" Again, the word outgroup was replaced with either *Wiconi students* or *Canug student*s depending on the school program of the participant. Participants answered using a four point response scale anchored by "I would very much not like to" (1) and "I would very much like to"

(4). See appendix N for the desire for future contact measure as it appeared to participants.

**Self-Assessment Manikin (SAM):** The SAM is used to evaluate stimuli or events on three domains: pleasure, arousal, and dominance (Bradley & Lang, 1994; Mizael et al., 2016). In this study, as in the Mizael et al. (2016) study, only the domain of pleasure was used. The SAM is an arrangement of five manikins, placed side by side, that progress from an unhappy and frowning manikin to a smiling manikin (see Figure 2 for an example of a SAM). Participants can indicate their selection by filling in the circle below or between the manikin that represents the pleasure they feel when presented with a specific stimuli. There are nine possible selections for any one stimulus. The circle farthest to the left, which represents the least pleasurable rating, has a numerical value of one. The opposite end of the scale on the far right, representing the most pleasurable rating, has a numerical value of nine. To start, participants were asked what feelings each manikin is displaying and corrected if wrong. This was done with the experimenter pointing to each manikin sequentially from left to right and asking the participant what feeling the selected manikin is showing.

Following the correct identification of the feelings shown by the manikins, the children were presented with the same manikin array with the addition of one experimental stimulus at a time, each presented in a randomly selected sequence. Participants rated all stimuli used in DMTS and testing (13 stimuli total) on the SAM. Participants were instructed to choose whichever circle (below and between the manikins) best represents what they feel when they looked at the stimuli (picture)

displayed on the top. Participants were told that there are no right or wrong answers as people have different tastes and preferences for things like clothes, movie characters, and so forth. See Appendix O for the SAM as it appeared to participants.

**Procedure**

***(Pre-test).*** Prior to DMTS training, all study participants were administered the following assessment measures (listed in the order in which they were presented):

1. The MEIM
2. Trait Ratings of Ingroup and Outgroup Members
3. Intergroup Attitude Measure
4. Ingroup and outgroup behavioral intentions measure
5. Desire for future contact measure
6. The SAM

Thirteen of the 15 participants completed these six measures in one session (approximately 25 minutes per participant) across different days prior to DMTS training and testing. Two participants completed all six of the measures in the same session as the DMTS training and testing. All participants completed DMTS training and testing within one session (approximately 60 minutes) and for all participants, the post-test measures were conducted immediately after completing the computerized DMTS training and testing.

Table 5 depicts the training and testing blocks that comprised the experiment. After completing all six initial measures, participants took the (DMTS) AC3 pre-test using the stimulus equivalence program on the school computers. During the pre-test,

each trial displayed a choice of three comparison stimuli to be matched with a presentation stimulus. The presentation stimulus was either the A1 (positive) stimulus or the A2 (negative) stimulus as a sample, accompanied by three comparison stimuli: a picture of an East African face (C1), an abstract symbol (C2), and a picture of a Native American face (C3). Each presentation of a C1 or C3 stimulus for comparison was randomly drawn from the bank of faces for the corresponding stimuli. The AC3 pre-test was composed of 16 trials, eight trials with the A1 stimulus and eight trials with the A2 stimulus, in a randomized sequence. There were no differential consequences for any selections. Participant responses were recorded by the computer program during this phase to see which stimuli participants chose to match with the positive and negative stimuli prior to any relations being directly trained.

**Pre-training.** A pre-training procedure was conducted to familiarize participants with an arbitrary delayed match to sample (DMTS) task. This training taught participants conditional relation X1Y1, X2Y2, and X3Y3. The stimuli used in this pre-training phase were pictures of animals (X1, X2, X3) and flowers (Y1, Y2, Y3) and were not used outside of this pre-training phase. Participants were instructed that they would be shown a first picture followed by a short delay (2 seconds) at which point they need to figure out which of the new pictures “goes with” the first picture. Participants were told that during the initial trials they would be presented with a cue indicating the correct choice. They were told to pay attention, as after the first attempts the cue would no longer be provided. On the first trial they were given only the sample stimulus as well as the correct comparison accompanied by two blank squares. On the second trial, they were given the

sample stimulus and two comparison stimuli as well as one blank square. On the third trial all three comparison stimuli were presented.

**Delayed match-to-sample trials.** Trials began with the appearance of the presentation stimulus. After a 2s interval, either 2 or 3 comparison stimuli (depending on the experimental phase) appeared. Differential consequences were presented in training blocks but not in testing blocks. Incorrect responses produced a tone (sharp buzzer) and a 2s delay to the presentation of the next stimulus. After each correct response, the computer program provided a tone (pleasant ding) to signal a correct response and tallied an earned point on the screen. Participants were given the following instructions:

> "You are going to do a computer program that asks you to match certain pictures with each other. When you make a correct match, you will earn a point. Points that you earn can be traded in for prizes when you are done with all the parts. If you do not score enough points in a section, you will get to try that section again and earn the maximum amount of points for it. These next stages will be similar to the block that you just completed (pre-training block) except that the pictures will be different and there will be no blank boxes to help you figure out the correct response. Instead you will need to pay attention to what happens when you make a choice such as listing for a sound or looking to see if you earned a point."

**Training.** Mastery criterion during training and testing was a maximum of one error per block. The participants were required to achieve mastery to advance to the subsequent block.

Table 5

*Stimulus Equivalence Training and Testing Schedule*

| Block | Number of Trials per Block | Mastery Criterion | Conditional Relations |
|---|---|---|---|
| AC3 Pretest | 16 | --------- | AC |
| Pretraining | 15 | 14/15 trials | X1Y1/X2Y2/X3Y3 |
| Training 1 AB 100% | 16 | 15/16 trials | A1B1/A2B2 |
| Training 2 AB 100% | 16 | 15/16 trials | A1B1/A2B2 |
| Test BA Symmetry | 16 | 15/16 trials | B1A1/B2A2 |
| Training 1 BC 100% | 16 | 15/16 trials | B1C1/B2C2 |
| Training 2 BC 100% | 16 | 15/16 trials | B1C1/B2C2 |
| Test CB Symmetry | 16 | 15/16 trials | C1B1/C2B2 |
| Mixed Training 1 ABBC 100% | 16 | 15/16 trials | A1B1/B2C2/B1C1/B2C2 |
| Mixed Training 2 ABBC 100% | 16 | 15/16 trials | A1B1/B2C2/B1C1/B2C2 |
| Baseline Review 1 50% | 16 | 15/16 trials | A1B1/B2C2/B1C1/B2C2 |
| Baseline Review 2 50% | 16 | 15/16 trials | A1B1/B2C2/B1C1/B2C2 |
| Test AC | 16 | 15/16 trials | A1C1/A2C2 |
| Test CA | 16 | 15/16 trials | C1A1/C2A2 |
| Test AC3 | 16 | ---------- | A1C1/A2C2 (with C3) |

Note. Training protocol adopted from Mizael et al. (2016)

After participants reached criterion in the pre-training phase, the AB relation was taught with a 16 trial block (eight A1B1 and eight A2B2). When criterion was met for both AB training blocks, a symmetry test (BA) followed and consisted of a 16 trial block

(eight B1A1 and eight B2A2). During symmetry tests participants were informed "This section is like the last two except you will not be able to see the points that you are earning or hear the sound that tells you if a choice was correct or incorrect. You will still be earning points that you can use for prizes when you are done, they are just hidden for this part. You must still get enough to move to the next part, if you do not get enough correct you will repeat the section until you do." When a participant achieved the criteria threshold in the BA symmetry test, they then proceeded to two successive BC training phases that followed the same procedure and structure as the AB training phases. CB symmetry tests were conducted after participants successfully met criterion for the two BC training phases. The next phases randomly mixed both AB and BC trials (A1B1/B2C2/B1C1/B2C2) and again required participants to correctly answer correctly on 15/16 trials on both mixed trial phases to proceed to the subsequent baseline review phase. The baseline review blocks of trials were nearly identical to the mixed training sequence, except that differential consequences were presented on only 50% of the trials. Similar to the mixed trials, baseline review was given in two successive blocks. Before this block began, participants were informed "in this next section you will sometimes see you points add up and hear sounds when you make your choices. Other times you will not see your points change or hear a sound. If you do not hear a sound or see your points it does not mean that you answered incorrectly. Even when you do not see your points they are still adding up for prizes when you are done."

**Equivalence tests**. Standard transitivity and equivalence tests were conducted by verifying AC and CA relations. During this test only C1 and C2 appeared as choices or

samples and each test block consisted of 16 AC and 16 CA trials in a randomized order. Participants needed to reach criterion in the AC test to proceed to the CA test, and then reach criterion again in the CA test to proceed to the final AC3 test. See Appendix P for a screenshot of the AC test, Appendix Q for a screenshot of the CA test, and Appendix R for a screenshot of the AC3 test.

**Posttests.** Following equivalence tests, participants were given the AC3 test in which AC trials displayed three comparison stimuli: an East African face (C1), an arbitrary symbol (C2), and a Native American face (C3). The AC3 test was run in a 16 trial block with no signals to the participants for correct or incorrect responses or a criterion that forced them to repeat the test. All participants completed the entire DMTS training and testing protocol in one session in the school computer lab, with most participants completing the entirety of the protocol in 50 to 60 minutes. Each participant was given a prize (e.g., hot Takis, gel pen, scented pen, or candy) of their choice following completion of all procedures. Participants were informed that the points they earned could be traded in for these prizes, but the criteria of correct responding set for each block meant that blocks would be repeated until criteria was met. Thus, each participant who completed the entire protocol earned the maximum amount of points, allowing all participants access to the prizes. After participants completed the AC3 test on the computer they were given the following 5 assessments, identical in form to the assessments completed prior to the stimulus equivalence training.

1. Trait Ratings of Ingroup and Outgroup Members
2. Intergroup Attitude Measure

3. Ingroup and Outgroup Behavioral Intentions Measure

4. Desire for Future Contact Measure

5. The SAM

# Chapter 4

# RESULTS

The primary research question guiding this study asked how the optimized training parameters from the Mizael et al. (2016) study could be extended and results replicated for the novel participant population in this study. Mizael et al. (2016) emphasized the implementation of their redesigned training protocol as a critical factor contributing to the improved results in their study compared to the de Carvalho and de Rose (2014) results. The study presented in this dissertation attempted to replicate the training protocol that proved so successful to Mizael et al.

## DMTS Training

A review of the training results in this study allows for a comparison to the training results in the Mizael et al. study to determine if participants progressed through the protocol in a similar manner. Differences in participant performance in training would potentially have implications for any differences found in testing.

Table 6 shows the DMTS training results for all 15 participants. Ten of the participants were able to complete the pre-training in one or two blocks. Two participants required three repetitions of the pre-training, two participants needed four attempts, and one participant took five attempts to meet criterion to proceed. The AB training phase

appears to show a faster acquisition than the pre-training phase with all of the participants reaching criterion in 3 or fewer trials. In the second round of AB training 14 of the 15 participants only needed one trial to reach criterion, with the 15th needing just two. The total number of trials required to reach criterion across all participants started at 33 during pre-training, 21 for the Training 1 AB block, and dropped to 16 in the Training 2 AB block. A similar pattern is seen with the introduction of the C stimuli in the training BC blocks. In the Training 1 BC block, three of the 15 participants needed more than two attempts to meet criterion, with one participant needing five attempts. After meeting criterion in Training 1 BC, all of the students were able to reach criterion in the Training 2 BC block with two or fewer attempts. Across all participants, the number of total attempts to reach criterion dropped from 27 in the Training 1 BC block to 17 in the Training 2 BC block. The first block of each new relation introduced and trained (pre-training, Training 1 AB, and Training 1 BC) required 1.8 attempts on average by participants to reach criterion. After participants were able to reach criterion in those blocks that introduced a new relation, the number of attempts dropped to an average of 1.1 attempts for each participant in the Training 2 AB and Training 2 BC blocks. In the first baseline review phase in which feedback was reduced to 50% of the trials, all 15 participants reached criterion in just one attempt. In the second baseline review phase, 13 of the 15 participants again needed just one attempt to achieve criterion. One student needed two attempts to reach criterion, and one participant needed nine attempts to reach the required level. Nine attempts was by far the highest number of attempts needed by any participant in any of the training phases. This participant reported feeling bored and

frustrated with the training. After the experimenter reminded them of the need to continue repeating until a high enough score was reached to "pass" and be finished, the participant reached criterion on their next attempt. On average, participants in the Mizael et al. study required 1.17 attempts to reach criterion in all four of the training AB and training BC blocks. Similarly, participants in the present study required an average of 1.35 attempts to reach criterion in these same blocks.

Table 6

*Participant Performance on DMTS Training Blocks*

| Participant ID | Pretraining | AB1 | AB2 | BC1 | BC2 | ABBC1 | ABBC2 | BL Review 1 | BL Review 2 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 3(15/15) | 2(16/16) | 1(16/16) | 1(15/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| 3 | 4(15/15) | 2(16/16) | 1(16/16) | 3(16/16) | 1(16/16) | 1(15/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| 4 | 5(15/15) | 1(16/16) | 1(16/16) | 2(16/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| 7 | 1(15/15) | 3(15/16) | 1(16/16) | 3(15/16) | 2(16/16) | 1(15/16) | 1(15/16) | 1(16/16) | 9(15/16) |
| 8 | 3(14/15) | 2(16/16) | 1(16/16) | 2(16/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(15/16) |
| 14 | 4(14/15) | 1(16/16) | 1(15/16) | 2(16/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(15/16) |
| 15 | 2(14/15) | 1(16/16) | 1(15/16) | 1(15/16) | 1(15/16) | 1(15/16) | 1(16/16) | 1(16/16) | 1(15/16) |
| 10 | 1(15/15) | 1(16/16) | 1(16/16) | 1(15/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(15/16) |
| 12 | 1(15/15) | 1(16/16) | 1(16/16) | 1(16/16) | 2(16/16) | 1(16/16) | 2(16/16) | 1(15/16) | 1(15/16) |
| 5 | 2(15/15) | 1(15/16) | 1(16/16) | 2(16/16) | 1(15/16) | 1(15/16) | 1(16/16) | 1(15/16) | 1(16/16) |
| 16 | 2(15/15) | 1(15/16) | 1(16/16) | 1(15/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| 13 | 1(4/15) | 1(15/16) | 2(16/16) | 2(15/16) | 1(16/16) | 1(15/16) | 1(15/16) | 1(16/16) | 2(15/16) |
| 2 | 1(15/15) | 3(16/16) | 1(15/16) | 1(15/16) | 1(16/16) | 1(16/16) | 1(15/16) | 1(15/16) | 1(15/16) |
| 19 | 1(15/15) | 1(15/16) | 1(16/16) | 5(16/16) | 1(15/16) | 1(16/16) | 1(15/16) | 1(16/16) | 1(15/16) |
| 23 | 2(15/15) | 2(16/16) | 1(16/16) | 1(15/16) | 1(15/16) | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| Total Trials | 33 | 21 | 16 | 27 | 17 | 15 | 16 | 15 | 24 |

**DMTS Testing**

Table 7 shows the results for equivalence tests. No feedback was given for responses in any of the testing phases. The pre-test AC3 column indicates the number of trials (out of 16 total trials) in which a participant matched the positive stimulus to an outgroup face or matched the negative stimulus to the abstract shape stimulus (C2). Eight

trials in the AC3 pre-test presented participants with the positive stimulus (A1) as the initial presentation stimulus with three stimuli as response options (ingroup face, abstract shape, and outgroup face). Eight trials presented participants with the negative stimulus (A2) with three stimuli as response options (ingroup face, abstract shape, outgroup face). Presentation of either the A1 or A2 stimuli was randomized by the computer program for participants. These responses were collected prior to directly training these relations so relatively low scores were expected.

In the BA symmetry block, 14 of the 15 participants reached criterion in one attempt. Similarly, in the test CB symmetry block 14 of the 15 participants met criterion in one attempt.

The AC test evaluated the property of transitivity (i.e. if A=B, and B=C, then A=C). All 15 study participants demonstrated the property of transitivity on either their first (12 participants) or second attempt (3 participants). The CA test evaluated the property of equivalence. Fourteen of the 15 participants were able to demonstrate the property of equivalence on their first attempt (13 participants) or second attempt (1 participant). One participant took the test nine times before demonstrating equivalence. Similar to the outlier in the training phase of nine attempts, this participant reported being frustrated and feeling "bored." When an experimenter redirected the participant by explaining they were close to being done and almost able to trade their points in for prizes, the participant promptly and correctly answered on all 16 trials in their ninth attempt at the test. In sum, all participants, with one exception, were able to demonstrate all the emergent relations tested in two or fewer attempts.

Table 7

*Participant Performance on DMTS Testing Blocks*

| Participant ID | Pretest (AC3) | BA Symmetry Test | CB Symmetry Test | AC Test | CA Test |
|---|---|---|---|---|---|
| 1 | 6 | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| 3 | 2 | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| 4 | 8 | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| 7 | 3 | 1(16/16) | 3(16/16) | 1(15/16) | 1(15/16) |
| 8 | 3 | 1(16/16) | 1(15/16) | 1(16/16) | 1(16/16) |
| 14 | 6 | 1(15/16) | 1(16/16) | 1(15/16) | 9(16/16) |
| 15 | 0 | 1(15/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| 23 | 12 | 1(16/16) | 1(16/16) | 1(15/16) | 1(15/16) |
| 2 | 6 | 1(16/16) | 1(16/16) | 1(16/16) | 1(15/16) |
| 5 | 10 | 1(15/16) | 1(15/16) | 2(15/16) | 1(16/16) |
| 10 | 1 | 1(16/16) | 1(16/16) | 2(16/16) | 1(15/16) |
| 12 | 9 | 1(16/16) | 1(16/16) | 1(16/16) | 1(16/16) |
| 13 | 2 | 1(15/16) | 1(16/16) | 1(16/16) | 2(15/16) |
| 16 | 4 | 1(16/16) | 1(16/16) | 2(16/16) | 1(15/16) |
| 19 | 5 | 3(15/16) | 1(16/16) | 1(15/16) | 1(15/16) |

Note: The number outside the parentheses indicates the number of attempts to reach criterion, the numbers inside the parentheses indicate the number of correct responses / the total possible in the final attempt in each block. The Pretest AC3 and final AC3 test were the total number of correct responses out of 16 trials.

The final AC3 test expanded the response options from consisting only of stimuli used in training blocks (outgroup face and abstract shape) to all three of the C stimuli (outgroup face, abstract shape, and ingroup face). Only one of the 15 participants was able to provide the correct response in 15 or more of the 16 trials in the AC3 post-test. Interestingly, it was the participant who took nine attempts to complete the CA test. It is possible that response relations were strengthened for this participant from the additional repetitions of the CA testing block. However, no feedback was provided during that testing block so it would seem unlikely that these extra repetitions would have provided much additional training above and beyond what other participants received. These

results are consistent with results in previous stimulus equivalence research with socially relevant stimuli (Watt et al., 1991; Carvalho & Rose, 2014; Haydu, Camargo, & Bayer, 2015).

**Measures of Stereotyping and Prejudice**

This study attempted to examine the potential generalization of stimulus class formation as indexed by measures of stereotypes and prejudice. Tables 8 through 15 contain the results from the MEIM, the trait ratings measure, the intergroup attitudes measure, the behavioral intentions measure, the desire for future contact measure, and the SAM. Results for each individual measure are provided in the accompanying sections. Taken together, the results from these measures provide little evidence for generalization of stimulus class formation beyond the experimental context. Potential reasons for this will be offered in the discussion section.

**MEIM.** Table 8 shows the mean responses from each ethnic group of participants. There is no threshold score for the MEIM that indicates when a group or individual is "committed" to their ethnic identity. Rather these scores allow for comparisons between groups of students. MEIM responses from participants in this study had a possible score range of one to four. Both the Native American and East African participants had similar ratings in the domain of ethnic exploration. In items related to the domain of ethnic group commitment, the East African participants had a slightly higher mean response. Both groups had overall mean responses above 3, with East African students approaching the maximum score. According to the MEIM, the subscores for the separate domains indicate that both of these student groups exhibited strong ethnic

identity in terms of both exploration of their ethnic identity and commitment to their ethnic identity through responses to items on the MEIM.

Table 8

*Mean Ratings on the MEIM*

| Ethnicity | Exploration | Commitment | Overall |
|---|---|---|---|
| Native American | 3.47 | 3.23 | 3.35 |
| East African | 3.58 | 3.87 | 3.72 |

**Trait Ratings.** Table 9 shows the mean responses on the trait ratings measure. Response scores for this measure ranged from 1 to 4, with positive traits reverse scored so that higher scores indicate a stronger endorsement of negative stereotypes. Positive traits were reverse scored, so higher scores indicate more negative stereotypes towards the target group. At the outset of the study, participants showed slightly greater negative stereotypes towards the outgroup (2.34) than the ingroup (1.78). Following stimulus equivalence training and testing, a shift was observed with the outgroup mean rating rising to 2.56 and the ingroup mean rating decreasing to 1.67. Breaking apart the participants into ethnic groups, the East African students mirrored this pattern with ingroup mean ratings dropping from 1.65 before the intervention to 1.37 following training and testing. The outgroup rating increased from 2.55 before training and testing to 2.86 following the intervention. The Native American participants ingroup ratings remained fairly consistent before (1.93) and after (2.01) the intervention. The outgroup ratings for these participants increased from 2.09 to 2.22.

Table 9

*Mean Trait Ratings Scores Before and After DMTS Training and Testing*

| | Ingroup Pre | Ingroup Post | Outgroup Pre | Outgroup Post |
|---|---|---|---|---|
| All Participants | 1.78 | 1.67 | 2.34 | 2.56 |
| Native American | 1.93 | 2.01 | 2.09 | 2.22 |
| East African | 1.65 | 1.37 | 2.55 | 2.86 |

Table 10 shows the trait ratings scores at the individual participant level. Mean participant ratings are shown for the ingroup and outgroup before and after the stimulus equivalence training and testing as well as the difference between outgroup and ingroup ratings before and after the intervention (mean outgroup rating minus the mean ingroup rating). This table provides a more detailed look at the range of responses from participants. Participants 1, 2, 5, 10, 12, 14, 16, and 19 attributed positive and negative attributes to the ingroup and outgroup relatively evenly (seen in the outgroup –ingroup columns). On the other end of the response spectrum, a number of participants showed dramatic differences in their ratings of the ingroup and outgroup. Participants 3, 4, 7, 13, and 15 had a mean difference between the ingroup and outgroup of 1.78 or more either before or after the training and testing. Brown et al. (2016; 2017) argue that a frequent mistake in using ratings such as these is in the interpretation of differences. A pattern of favoring one's ingroup does not represent the act of stereotyping a group. Instead, prejudice is an expression of negativity toward an individual based on their group membership. These five participants (3, 4, 7, 13, and 15) provide an example of this pattern. Not only do they favor the ingroup in the trait ratings measure, the scores towards the outgroup approach the maximum range for the measure. The scale of responses for this measure included the options none (1), some (2), most (3), and all (4).

Table 10

*Participant Mean Scores for Ingroup and Outgroup*

| Participant Number | Ethnicity | Ingroup Avg Pre | Ingroup Ave Post | Outgroup Avg Pre | Outgroup Avg Post | Outgroup - Ingroup Pre | Outgroup-Ingroup Post |
|---|---|---|---|---|---|---|---|
| 1 | East African | 1.44 | 1.56 | 1.44 | 1.56 | 0.00 | 0.00 |
| 3 | East African | 1.22 | 1.00 | 2.89 | 3.67 | 1.67 | 2.67 |
| 4 | East African | 1.67 | 1.11 | 3.00 | 3.44 | 1.33 | 2.33 |
| 7 | East African | 1.11 | 1.00 | 3.44 | 4.00 | 2.33 | 3.00 |
| 8 | East African | 1.78 | 1.89 | 2.56 | 3.11 | 0.78 | 1.22 |
| 14 | East African | 2.22 | 2.22 | 2.11 | 2.11 | -0.11 | -0.11 |
| 15 | East African | 1.67 | 1.22 | 4.00 | 4.00 | 2.33 | 2.78 |
| 23 | East African | 2.11 | 1.00 | 1.00 | 1.00 | -1.11 | 0.00 |
| 2 | Native American | 1.33 | 2.00 | 1.33 | 2.00 | 0.00 | 0.00 |
| 5 | Native American | 1.89 | 2.33 | 1.56 | 2.44 | -0.33 | 0.11 |
| 10 | Native American | 2.11 | 1.89 | 2.11 | 1.78 | 0.00 | -0.11 |
| 12 | Native American | 1.78 | 2.22 | 1.56 | 2.11 | -0.22 | -0.11 |
| 13 | Native American | 2.22 | 1.78 | 4.00 | 3.56 | 1.78 | 1.78 |
| 16 | Native American | 2.11 | 1.78 | 2.00 | 1.89 | -0.11 | 0.11 |
| 19 | Native American | 2.11 | 2.11 | 2.11 | 1.78 | 0.00 | -0.33 |

Participants 3, 4, 7, 13, and 15 all had outgroup mean ratings of at least 3.44 on the post stimulus equivalence administration of the trait ratings measure. They consistently responded with most or all when asked how many of the students in the other

program were mean, unfriendly, selfish, and naughty. They also consistently responded with none or some when asked how many in students in the other school program were nice, friendly, helpful, smart, or good. When combined with their lower mean scores for traits towards their ingroup, these participants not only produced responses that favored their ingroup, but demonstrated negative trait attributions towards students in the other school program based on their group membership.

**Intergroup Attitudes**: Table 11 shows the mean response scores for the intergroup attitude measure. Response scores for this measure ranged from 1 (very untrue of what I believe) to 6 (very true of what I believe). Lower scores indicate more negative attitudes towards the outgroup. The mean response for all participants was stable at 4.07 prior to stimulus equivalence training and testing and 4.02 after. When looking at the individual response scores, a similar pattern is seen to the previous trait ratings measure. Participants 3, 4, 7, 15, and 13 had five of the six lowest scores on the measure.

Table 11

*Mean Scores on the Intergroup Attitude Measure*

| Participant | Ethnicity | Mean Pre | Mean Post |
|---|---|---|---|
| ALL | Both Groups | 4.07 | 4.02 |
| 1 | East African | 5.45 | 5.45 |
| **3** | **East African** | **3.45** | **2.64** |
| **4** | **East African** | **3.91** | **2.09** |
| **7** | **East African** | **3.00** | **3.73** |
| 8 | East African | 4.82 | 3.30 |

| | | | |
|---|---|---|---|
| 14 | East African | 4.36 | 4.64 |
| **15** | **East African** | **1.55** | **2.60** |
| 23 | East African | 4.91 | 5.55 |
| 2 | Native American | 6.00 | 4.45 |
| 5 | Native American | 4.64 | 3.45 |
| 10 | Native American | 3.55 | 4.64 |
| 12 | Native American | 4.82 | 5.00 |
| **13** | **Native American** | **1.30** | **3.73** |
| 16 | Native American | 4.73 | 5.36 |
| 19 | Native American | 4.64 | 3.73 |

Note: Participants in bold were the five participants who demonstrated the strongest endorsement of negative stereotypes towards the outgroup on the trait ratings measure. Mean scores from these participants are put in bold to aid in the comparison of their responses on additional measures compared to other study participants.

**Behavioral Intentions.** Participants were shown a picture of either an ingroup or outgroup student and presented with the hypothetical scenario that this person arrived as a new student in their classroom. They were then asked to respond to questions such as, "how much would you like to eat lunch with this person?" Response options included very unhappy (1), unhappy (2), happy (3), and very happy (4). Table 12 shows the mean scores for participant responses. Across all participants mean scores towards the ingroup remained close, with a mean of 3.58 prior to stimulus equivalence training and testing and 3.5 after. Mean scores for the outgroup increased from a mean of 3.13 to 3.28 following the intervention.

Table 12

*Mean Scores on the Behavioral Intentions Measure*

| Participant Number | Ethnicity | Ingroup Mean Pre | Ingroup Mean Post | Outgroup Mean Pre | Outgroup Mean Post | Difference I/O Pre | Difference I/O Post |
|---|---|---|---|---|---|---|---|
| ALL | Both | 3.58 | 3.5 | 3.13 | 3.28 | 0.45 | 0.21 |
| 1 | East African | 4 | 4 | 4 | 4 | 0 | 0 |
| **3** | **East African** | **4** | **4** | **3** | **3** | **1** | **1** |
| **4** | **East African** | **2.75** | **2.5** | **2.5** | **3.5** | **0.25** | **-1** |
| **7** | **East African** | **4** | **4** | **3** | **2.5** | **1** | **1.5** |
| 8 | East African | 4 | 2 | 4 | 3 | 0 | -1 |
| 14 | East African | 4 | 4 | 3 | 3.75 | 1 | 0.25 |
| **15** | **East African** | **3.75** | **2.75** | **2** | **2.5** | **1.75** | **0.25** |
| 23 | East African | 3 | 4 | 3.25 | 4 | -0.25 | 0 |
| 2 | Native American | 4 | 4 | 4 | 3.75 | 0 | 0.25 |
| 5 | Native American | 4 | 2.75 | 4 | 2.25 | 0 | 0.5 |
| 10 | Native American | 3 | 3.5 | 2.75 | 3.25 | 0.25 | 0.25 |
| 12 | Native American | 3.5 | 3.25 | 3.25 | 3 | 0.25 | 0.25 |
| **13** | **Native American** | **3** | **4** | **1** | **3.25** | **2** | **0.75** |
| 16 | Native American | 3.5 | 4 | 3.25 | 4 | 0.25 | 0 |
| 19 | Native American | 3.25 | 3.75 | 4 | 3.5 | -0.75 | 0.25 |

Note: Participants in bold were the five participants who demonstrated the strongest endorsement of negative stereotypes towards the outgroup on the trait ratings measure. Mean scores from these participants are put in bold to aid in the comparison of their responses on additional measures compared to other study participants.

Table 12 also includes individual participant mean scores for the measure, with the five students in bold who had the most divergent ingroup and outgroup ratings on the

trait ratings measure. Across all participants the mean difference in ratings between the ingroup and outgroup before DMTS trainging and testing was 0.45, which decreased to 0.21 in the post-DMTS administration of the behavioral intentions measure. The mean rating for this measure from the group of five participants (3, 4, 7, 13 and 15) highlighted from the previous trait ratings measure had a mean of 1.2 prior to DMTS training and testing. Interestingly this group showed a larger reduction in ingroup/outgroup ratings differences compared to all participants on the post-DMTS administration of the behavioral intentions measure with the mean decreasing to 0.5.

**Desire for future contact.** Participants responded to three items to measure their desire or willingness to engage in future contact with outgroup members. Responses ranged from "I would very much not like to" (1) to "I would very much like to" (4). Table 13 shows the mean responses for the measure before and after stimulus equivalence training and testing. There was a decrease in the mean response score across all participants from 3.07 to 2.49. This was likely due to two participants (5 and 7) dramatically lowing their ratings from one administration of the measure to the next. As with the other measures, the five participants who had the most negative trait ratings towards the outgroup are in bold in the table. As observed in the other measures, a similar pattern of responding was apparent on this measure, with those five participants having the lowest mean ratings on the second administration of this measure.

Table 13

*Mean Scores for the Desire For Future Contact Measure*

| Participant Number | Ethnicity | Mean Pre | Mean Post |
|---|---|---|---|
| ALL | Both | 3.07 | 2.49 |
| 1 | East African | 3.00 | 3.00 |
| **3** | **East African** | **2.33** | **1.33** |
| **4** | **East African** | **3.67** | **1.67** |
| **7** | **East African** | **4.00** | **1.00** |
| 8 | East African | 3.00 | 2.00 |
| 14 | East African | 3.00 | 3.67 |
| **15** | **East African** | **1.00** | **1.00** |
| 23 | East African | 3.00 | 4.00 |
| 2 | Native American | 3.67 | 4.00 |
| 5 | Native American | 4.00 | 1.67 |
| 10 | Native American | 3.33 | 2.67 |
| 12 | Native American | 4.00 | 3.00 |
| **13** | **Native American** | **1.33** | **1.00** |
| 16 | Native American | 3.33 | 4.00 |
| 19 | Native American | 3.33 | 3.33 |

Note: Participants in bold were the five participants who demonstrated the strongest endorsement of negative stereotypes towards the outgroup on the trait ratings measure. Mean scores from these participants are put in bold to aid in the comparison of their responses on additional measures compared to other study participants.

**SAM.** Table 14 shows the group mean stimulus ratings for five stimuli presented by the SAM (A1, A2, B1, B2, C2) before and after DMTS training and testing (across all

participants). Refer to Figure 3 for pictures of stimuli used. The SAM asked participants to rate all stimuli used in DMTS training and testing in the domain of pleasure. Participants were shown the stimuli one at a time and instructed, "Please rate the picture I am showing you based on how much pleasure you feel when you look at it." Participants were shown an array of manikins anchored by a frowning and unhappy manikin on the left to a smiling manikin on the right. Score responses ranged from 1 (very unpleasant/frowning manikin) to 9 (very pleasant/smiling manikin). Although all stimuli used in the study were rated by all of the participants, only five stimuli (A1, A2, B1, B2, and C2) were used in DMTS training by both groups of students. The other eight stimuli were pictures of ingroup and outgroup faces. Only outgroup faces were used as stimuli in the DMTS training procedures, and the faces used as the outgroup were dependent on the ethnicity of the participant. Higher scores indicate greater reported levels of pleasure by participants. Stimulus A1, the smiling emoji with the thumbs up, was the most highly rated of any stimuli in the study. The high ratings given to this stimulus help to corroborate the stimulus pilot selection of this stimulus as one that they would "pair or match with a person that you like, that you would want to spend time with, that you think of as a friend." Across all participants, this stimulus increased slightly in mean rating from 8.27 before the intervention to 8.6 following it. The A2 stimulus was the lowest rated stimulus in the study, also corroborating the selection of this stimulus by the stimulus selection pilots to select an image that they would "pair with a person that you do not like, that you would not want to spend time with, that you think of as unfriendly." This stimulus also increased in mean rating from 2.07 at pretest to 2.8 at posttest. The

SAM ratings for the abstract shapes, B1, B2, and C2 prior to stimulus equivalence training and testing ranged from 5.73 to 6.2. Most participants reported feeling unsure about what the shape was or how they should rate it, which of course was the purpose of using abstract shapes with no pre-experimental meaning to the participants. Following training and testing, mean SAM scores for these stimuli changed, a change which would be expected to occur if there was a transfer of function between a formed class of stimuli. During training, participants earned reinforcement for matching the B1 stimulus with the positive stimulus A1. Given that participants indeed viewed stimulus A1 as positive, as demonstrated by their SAM ratings, successful formation as members in a class of stimuli should transfer some degree of that positive valiance to the other stimuli in the class. Stimuli that formed a class with the positive stimulus (A1) should have higher SAM ratings following DMTS training and testing. Conversely, stimuli that formed a class with the negative stimulus (A2) should have lower SAM ratings following training and testing). The data in part support this, as the mean rating for stimulus B1 increased from 6.14 to 7.07. Similarly, the B2 and C2 stimuli were successfully grouped in a class with the negative A2 stimulus. The mean rating for the B2 stimulus decreased from 6.2 to 4.47 and the C2 stimulus decreased from 5.73 to 4.87. The magnitude of change in the mean rating for these stimuli is difficult to contextualize given the small number of participants, but that each of them changed in the expected direction based on their pairing with either the positive or negative stimulus suggests some degree of influence was exerted on these stimuli based on their formation in a class with either the positive stimulus or negative stimulus.

Table 14

*Mean Stimulus Ratings Pre and Post*

| Stimulus | Mean Pre | Mean Post |
|---|---|---|
| A1 | 8.27 | 8.60 |
| A2 | 2.07 | 2.80 |
| B1 | 6.14 | 7.07 |
| B2 | 6.20 | 4.47 |
| C2 | 5.73 | 4.87 |

Table 15 shows the mean SAM ratings across participants for outgroup faces. For each group of participants, a different set of faces served as the outgroup face stimuli. Prior to DMTS training and testing, participants had a mean SAM rating towards outgroup faces of 5.63. Following DMTS training and testing the mean SAM rating for outgroup faces increased to 6.54. Higher ratings on the SAM (rating scale 1-9) indicate greater reported "pleasure" by participants towards a stimulus.

Table 15

*Mean SAM Ratings for Outgroup Face Stimuli*

| Outgroup Faces Pre | Outgroup Post |
|---|---|
| 5.63 | 6.54 |

## CHAPTER 5

The primary research question in this study asked whether the training parameters used by Mizael et al. (2016) could be extended and results reproduced for a novel group of participants. The test results shown for the AC and CA tests indicate that this question was answered in the affirmative. This study used the "optimized" training protocol from previous stimulus equivalence and bias research done by Mizael et al. (2016). The participants were students in a segregated school environment that has struggled with conflict between populations of East African and Native American students separated by program in the same physical building for years. Fourteen of the 15 study participants were able to show equivalence class formation of a positive stimulus and stimuli of faces representing the outgroup relevant in this particular school setting.

De Carvalho and de Rose (2014) were the first to apply the conflicting relations paradigm, equivalence relations that contradict the pre-experimental histories of the participants, to racial bias. In this initial study, only one of four participants was able to demonstrate class formation with stimuli that included a positive symbol and black faces. Noting that this result was a typical outcome with research in the conflicting relations paradigm, Mizael et al. (2016) expanded the training protocol and found that all 13 participants in their study were able to demonstrate class formation, a notable achievement in this tangential line of stimulus equivalence research. The study that is the subject of this paper was a systemic replication of the methods that proved so successful for the Mizael et al. study, with a novel group of participants. Results of the present investigation suggest that standard equivalence post-tests confirmed the effectiveness of

this training protocol in successfully forming stimulus classes using stimuli that had pre-experimental meaning to study participants.

**Identification of Bias**

One difference between this study and the previous stimulus equivalence bias research was the selection of participants. Both de Carvalho and de Rose (2014) and Mizael et al. (2016) screened potential participants and selected only those that demonstrated bias on their chosen measures participated in the study. The screening procedure used in these previous studies consisted of a DMTS task in which participants were given either the positive or negative stimulus followed with the response option of pairing that stimulus with either a white face, abstract shape, or black face. A bias value (*b*) was derived from [(white face to positive – black face to positive) – (black face to negative – white face to negative)]. Higher values indicated a negative bias towards black faces, and enrollment in the study was based on having a *b* value equal or greater than four. Although the present study did not utilize this inclusion criteria, an identical measure (the AC3 pre-test) was given prior to the start of the training protocol. Table 16 shows response by response results for each participant on the pre-test. Using the same formula as Mizael et al. (2016), with white and black swapped out for ingroup and outgroup, four participants met or exceeded Mizael et al.'s threshold of $b >= 4$ indicating a negative bias. Specifically, participants 2, 3, 10, and 19 all had a value of *b* at or greater than 4, suggesting they were most biased against the outgroup. However this is just one lens through which we examined. All four of those participants were able to demonstrate

class formation in the equivalence tests, a result that directly replicates the results in the Mizael et al. (2016) study.

Table 16

*Participant Pre-test Responses*

| Participant Number | Positive Stimulus | Negative Stimulus | *b* |
|---|---|---|---|
| 1 | IOOIIOII | O | 2 |
| 2 | IOOIIIII | OOOI | 6 |
| 3 | O | OOOOO | 4 |
| 4 | | O | 1 |
| 5 | OOO | IOOOOOI | 0 |
| 7 | OIOOIO | IOIIOI | -4 |
| 8 | OOOOOO | IOOOOOI | -8 |
| 10 | IIIIII | OOOOOOO | 13 |
| 12 | IOOOOOOO | IIOIIO | -8 |
| 13 | IIIIIOI | IOIIIIO | -2 |
| 14 | OOIOOI | IIII | -6 |
| 15 | OOOOOOOO | OOOOIIII | -8 |
| 16 | IOI | IIIOI | -2 |
| 19 | IIIIII | OOO | 10 |
| 23 | OOIIOIII | I | 1 |

Table 16. Participant pre-test AC3 responses. Results from this test were presented previously in Table 10. This table shows individual responses from this test. When given the positive or negative stimulus, participants could match with the ingroup stimulus (I), the outgroup stimulus (O), or the abstract shape (not-coded). The value of *b* was calculated using the formula b=((I+) – (O+)) – ((O-) – (I-)). Responses in which participant matched the arbitrary stimulus (C2) with either the positive or negative stimulus were not coded, hence the blank cell for participant 4 who only matched C2 with the positive stimulus.

The stereotype and prejudice measures included in this study allow for a closer examination of the attitudes participants hold towards outgroup members. Looking at the Trait Ratings measure, participants 3, 4, 7, 13, and 15 all had mean outgroup ratings of 3 or higher, with participants 13 and 15 having mean outgroup ratings of 4, the highest possible value. Higher values on this ratings measure indicate more negative stereotypes towards outgroup members. Participants 13 and 15 responded that they felt all of the

Native American students in the other program were mean, unfriendly, selfish, and naughty. They also responded "none" when asked how many Native American students were helpful, nice, friendly, smart and good. One of these participants asked an experimenter if their teachers would find out about their answers. When the experimenter told them that the information was confidential, the participant said that she did not like Native American students because when she was a kindergarten student they teased her, pulled at her hijab, and threw pencils at her. That participant is now a fifth grade student, and her present attitudes appear to be based (to some degree) on a negative intergroup interaction that happened six years ago. Participants 13 and 15 were explicit about their dislike for students from the other cultural group, and their trait ratings reflected their stated feelings. Yet the pretest AC3 ***b* value** for each of the participants was quite low, with $b = -2$ for participant 13 and $b = -8$ for participant 15. Neither participant would have met the inclusion criteria for the Mizael et al. (2016) study, whereas the additional stereotyping and prejudice measures in the present study identify them as holding perhaps the most negative attitudes of any of the participants towards outgroup. Conversely, three of the participants (2, 10, 19) that had high *b* values in the pretest assigned traits evenly between the ingroup and outgroup in both the pre and post DMTS administration of the Trait Ratings measure, showing an absence of stereotyping towards outgroup members. This response pattern was repeated across all of the stereotyping and prejudice measures, with participants 3, 4, 7, 13, and 15 consistently holding the most negative attitudes of all the study participants towards the outgroup, or responding with the least inclination to interact with the outgroup in both the pre and post DMTS stereotyping and prejudice

measures. The discrepancy between the bias measure used by Mizael at al. for screening and the participant responses on the measures used in this study suggest that attempting to identify racially biased participants for interventions should rely on multiple measures. Had a participant in the Mizael et al. study consistently matched white faces to the positive stimulus and only the abstract shape to the negative stimulus, they would have exceeded the *b* threshold and been identified as biased. If we return to the idea that prejudice is not only favoring one's ingroup but holding negative views towards people based on their membership in a group, more complete measures designed to capture this dynamic must be employed.

**Results of the AC3 Test**

De Carvalho and de Rose (2014) used the AC3 test to confirm stimulus class formation. Only one of their four participants was able to do so. Mizael et al. (2016) instead used the AC and CA tests to confirm stimulus class formation, and included the AC3 test as a second and "more stringent" test of responding. Nine of the 13 participants maintained relations between black faces and positive symbols. The results of the AC3 test in the present study is the only aspect in which the results of Mizael et al. were not replicated. Specifically, in the present study, one of the 15 participants maintained relations between outgroup faces and the positive stimulus in the AC3 test. Mizael et al. suggested that transitioning from training blocks to a test that presented stimulus response options not present during training may have compromised the ability of participants to demonstrate class formation. In the abstract of the Mizael et al. paper, the authors describe the final AC3 test as "a different and more stringent test" to examine whether

participants maintained relations between black faces and the positive stimulus. On one hand, the authors are correct that the AC3 is a more stringent test of responding. Participants would need to continue to match the black face to the positive stimulus when the response options are expanded to include stimuli not present during training. The fact that these additional stimuli were white faces muddies the water so to speak. Any number of additional stimuli could have been introduced in the AC3 test as a way to make it more "stringent." In the present study, only one of the 15 participants continued to respond with the expected relations in the AC3 test. This required that participant to select an outgroup face when presented with the positive stimulus, when an ingroup face was also present as a response option. One could argue that a more desired pattern of responding for participants would be to select either an ingroup or outgroup face when presented with the positive stimulus, and more importantly to not choose to match an outgroup face with the negative stimulus. Both the de Carvalho and de Rose and the Mizael et al. studies very wisely constructed their stimulus classes to avoid pairing any of the face stimuli with the negative stimulus. I would argue that another smart design choice would be to use an outcome measure that would focus primarily on the reduction of participants matching outgroup faces with the negative stimulus. To illustrate and focus on matching the negative stimulus to outgroup faces, Table 17 shows the frequency of this occurrence for all of the study participants in the pre and post AC3 test.

Table 17

*Number of Pairings of the Negative Stimulus and an Outgroup Face.*

| Participant Number | Outgroup to Negative Pre | Outgroup to Negative Post |
|---|---|---|
| 1 | 1 | 0 |
| **3** | 5 | 4 |
| **4** | 1 | 2 |
| **7** | 2 | 0 |
| 8 | 3 | 1 |
| 14 | 0 | 0 |
| **15** | 4 | 0 |
| 23 | 0 | 1 |
| 2 | 3 | 0 |
| 5 | 5 | 1 |
| 10 | 7 | 0 |
| 12 | 2 | 1 |
| **13** | 2 | 0 |
| 16 | 1 | 0 |
| 19 | 3 | 0 |
| Total | 39 | 10 |

During the pretest, participants matched the negative stimulus with an outgroup face 39 times. Following stimulus equivalence training and testing, that number dropped to 10. Nine of the 15 participants never matched the negative stimulus to an outgroup face in the final AC3 test, and four participants did so only once. This perhaps represents a point of conflict between the behavioral process that is stimulus equivalence and the application of this technology to stereotyping and prejudice. From a purely behavioral perspective, the AC3 test is more stringent, testing the strength of the trained relations with increased response options for participants. From the perspective of stereotyping and prejudice, this test is somewhat lacking. Matching either of the faces with the positive stimulus is a desirable response. The more problematic response matching of an outgroup

face to the negative stimulus. An outcome measure that is as equally concerned with the former as the later does not seem to be in alignment with the goal of reducing racial stereotyping and prejudice.

**Generalization**

The lack of generalization of stimulus class formation to the measures of stereotypes and prejudice was disappointing but not altogether surprising. It is possible that these results speak to an issue of dosage. The optimized training protocol taken from Mizael et al. (2016) and used in this study proved effective at promoting stimulus class formation and reducing the number of instances of participants pairing an outgroup face with the negative stimulus in the AC3 post test. As shown in chapter 2, the Mizael et al. training protocol is considerably more extensive than predecessors used in this small line of research. Despite the increase in number of training blocks and modifications regarding the placement of testing blocks, the dosage of this intervention should be considered to be relatively small. Most participants were able to complete the entire training and testing protocol in less than one hour of time. Would obtained results seen on the stereotyping and prejudice measures have been different with a change to the dosage? Here dosage could refer to the extension and expansion of training blocks over a number of days or weeks. Some of the participants in this study have had their attitudes and views towards the outgroup formed over a period of years, often through intergroup interactions that have produced tremendous conflict. It is possible that the vision of de Carvalho and de Rose (2014) and Mizael et al. for stimulus equivalence to shape more desirable attitudes in matters of race would need a much larger dosage to begin to generalize

beyond the experimental context of these studies. Future research might examine the effect of expanding the training schedule in such a way that participants are exposed to the relational training not simply in one sitting, but over a number of days or weeks. In addition, the question remains what effect this type of training protocol may have on interpersonal interactions among individuals or groups. This question has yet to be tested.

**Limitations**

There are several limitations to the present study. Primary among them was the number of participants in the study. The initial goal for the study was to enroll 35-60 participants. The restriction to only recruit participants from the after school program ultimately limited the number of possible participants, a limitation that was further exacerbated by staffing issues that capped the number of students in the after school program to a small fraction of the typical enrollment in the program. Although 15 participants was sufficient to address the primary research question, further examination of the stereotyping and prejudice measures was adversely affected by this lower number. All of the measures used in this study have been previously used in research and found to be reliable, well-constructed measures. These measures have not, to my knowledge, been used with East African and Native American students. A larger number of participants would have allowed for a comprehensive analysis of the properties of these measures to verify their appropriateness for these specific populations.

The way in which the data from the stereotyping and prejudice measures is displayed in this paper will be problematic to some. The scales used in these measures is ordinal, but presented as if it is interval or ratio. The small number of participants in the

study precluded the use of some statistical treatments that might have been used to address this. Different ways of presenting the data (such as percentages or frequency) were considered but due to the large amount of measures and items in the data it was decided that using means, although problematic, were the most clear way to communicate the responses gathered from these instruments. The means used in the display of data in this paper are intended to allow for the examination of differences of categorical responding, rather than making arithmetic assumptions that would be inappropriate for this data. For example, a mean trait rating towards the ingroup of one compared to a mean outgroup trait rating of four is intended to highlight the categorical difference of trait allocations between the ingroup and outgroup.

Access to further school information pertaining to study participants may also have been beneficial. Although it would be impossible to quantify the problematic intergroup interactions that these students have experienced, it may have proved valuable to be able to access school incident reports involving the participants in this study to look for documented incidents of conflict between the student groups.

A final limitation is that data were not collected to examine the degree to which trained responses maintained over time. Understanding how long the effects of training last on participant responding may provide useful information about the strength of the relations between stimuli that were trained.

**Conclusions**

Knowing the effects of stereotyping and prejudice on children and adolescents naturally necessitates the ability to promote effective interventions to address the

formation or remittance of existing stereotypes and prejudice. One long running strategy comes from the intergroup contact hypothesis (Allport, 1954) which postulates that contact between social groups can reduce existing negative intergroup attitudes. Allport (1954) asserted that the effect of this intergroup contact was greatest when four conditions were met; the groups had equal status, engaged in cooperation, shared common goals, and were provided with institutional support. In their 2006 meta-analysis on the research, Pettigrew and Tropp found that intergroup contact under these conditions was indeed effective at reducing prejudice.

Meeting all four of Allports optimal conditions for group contact provides significant challenges for those working in school settings. Often there can be reticence on the part of students to interact with outgroup members or logistical challenges to bring groups together (Paluck & Green, 2009). One possible alternative to direct contact has been the emergence of interventions that rely on *indirect* (sometimes referred to as *extended*) forms of contact. Vezzali, Stathi, and Giovannini (2012) conducted an intervention to examine the effects of book reading that focused on intercultural themes between adolescents. Participants assigned to the intercultural reading condition showed reductions in stereotyping, improved intergroup attitudes, an increased desire to engage in future contact with outgroup members, and more positive behavioral intentions towards the outgroup compared to participants assigned to groups that either did not read or read books not centered around intercultural themes. Cameron and Rutland (2006) used an indirect (extended) contact intervention to examine the impact on prejudice toward disabled students. Their intervention consisted of reading a series of stories to

children (n=67, aged 5-10 years) that centered on the friendships and adventures between disabled and non-disabled characters. For children in the main experimental group, stories were followed with a post-story discussion that heightened the salience of the group membership of the characters as disabled or non-disabled. To measure the attitudes of the children the authors used a trait rating measure and a desire for future contact measure similar to the measures used in this stimulus equivalence study. Results from the Cameron and Rutland study found that the indirect contact intervention led to increased positivity towards the disabled. Liebkind and McAlister (1999) used a similar indirect contact intervention with a large group (n=1480) of 13-15 year old students in Finland to study the promotion of tolerance. Students in schools that were assigned to the experimental group were given stories of friendship between ingroup and outgroup members as examples of successful intergroup contact. To measure tolerance toward foreigners, 17 items were adapted from a variety of sources to study prejudiced attitudes, behavioral intentions, and opinions. This measure was adopted and adapted as the intergroup attitudes measure in the present stimulus equivalence study. The authors found that intergroup tolerance held constant or increased for students in the experimental condition while attitudes decreased or remained constant for students in the control schools. The authors felt that the intervention produced significant changes but were skeptical that the type of changes seen in their measure represented an enduring or persistent change to the attitudes of students towards foreign students. Instead, they viewed the utility of an indirect contact intervention as having the potential to facilitate

successful future direct intergroup contact by influencing initial expectations for the contact and increasing positive rather than negative affect for direct contact situations.

Mizael et al. (2016) articulated a hope that the procedures used in their stimulus equivalence study might have promise as a component in behavioral interventions used to address racial bias. The DMTS procedures used in this study, and in previous conflicting relations paradigm studies, do not constitute a fully formed intervention ready to be used to address the significant issues of stereotyping and prejudice. Consideration of these procedures as or part of a future intervention should be considered relative to their place on a continuum of translational research. The present study, as well as the Mizael et al. study, should be considered near the basic research end of the continuum. To the best of my knowledge, these are the only two studies in the conflicting relations paradigm that have demonstrated successful equivalence class formation using social relevant stimuli. Questions of central importance, such as dosage, maintenance, and generalization have yet to be addressed through research. To that end, the potential of the procedures used in these studies as an intervention is unknown. If these procedures were to be utilized in the future in some capacity as an intervention they would most closely align with the work being done with indirect or extended contact interventions. Participants in the present study related problematic past interactions with outgroup members and expressed reticence to engage in future intergroup interactions. Pettigrew and Tropp (2006) highlighted the role of reducing intergroup anxiety, feelings of threat and uncertainty surrounding intergroup interactions, in paving the way for more positive intergroup interactions that can ultimately reduce prejudice. This might prove vital for groups who

have a history of active conflict (e.g., Brazilian soccer fans, Catholics and Protestants in Northern Ireland). A potentially effective intervention package designed to reduce or remit stereotyping and prejudice might commence with indirect contact interventions, perhaps including procedures similar to those used in this and the Mizael et al. study, before transitioning to direct intergroup contact situations that meet as many of the Allport (1954) criteria as is possible in a school setting or perhaps simultaneously for groups that have some prejudice that is not necessarily based in active conflict.

This study successfully replicated and extended the results of the Mizael et al. (2016) study with a novel population and with additional instruments and measures to examine stereotyping and prejudice. Given the recent "replication crisis" in psychology, the results of the present study to replicate the results found in the Mizael et al. study should be seen in a positive light. Particularly as the procedures used in the Mizael et al. study were applied to an entirely different social context, in a different country, with a different group of participants. Mizael et al. suggested that future research should use faces of children rather than adults as training stimuli, which this study did, and that the generalization of the effects should be assessed.

This study attempted to examine generalization, and found little substantive evidence there was any. It is possible that this could be an issue of dosage, and that future research that increases the quantity or the timespan of the procedures used in this study may find evidence that supports further generalization beyond the experimental context. Additional studies are needed to explore the issue of maintenance and to explore the potential of the stimulus equivalence procedures as part of an intervention package that

promotes positive intergroup interactions. The enhanced training protocol used in the Mizael et al. (2016) study and the present study appear to be appropriate for the creation of new stimulus classes. The importance of these newly formed classes is an important question that now needs to be addressed. Problematic intergroup interactions can form relations that lead to bias, stereotyping and prejudice. The ability to use DTMS procedures to form new equivalence classes does not necessarily mean unforming the relations that have developed through various life experiences for people. Can successful class formation actually change the way in which people interact with each other? This study builds on the work from Mizael at al. to provide evidence that equivalence classes utilizing socially relevant stimuli can be formed, possibly in a variety of social contexts. The larger question of the importance of these equivalence classes now needs to be addressed.

# References

Aboud, F., Mendelson, M., & Purdy, K. (2003). Cross-race peer relations and friendship quality. *International Journal of Behavioral Development*, *27*(2), 165–173. https://doi.org/10.1080/01650250244000164

Allport, G. (1954). *The Nature of Prejudice*. Oxford, England: Addison-Wesley.

Annett, J., & Leslie, J. (1995). Stimulus Equivalence Classes Involving Olfactory Stimuli. *Psychological Record*, *45*(3), 439–452.

Arntzen, E. (2006). Delayed Matching to Sample: Probability of Responding in Accord with Equivalence as a Function of Different Delays. *The Psychological Record*, *56*(1), 135–167. https://doi.org/10.1007/BF03395541

Arntzen, E., Grondahl, T., & Eilifsen, C. (2011). The Effects of Different Training Structures in the Establishment of Conditional Discriminations and Subsequent Performance on Tests for Stimulus Equivalence. *The Psychological Record*, *60*(3). Retrieved from http://opensiuc.lib.siu.edu/tpr/vol60/iss3/4

Bachay, J. (1998). Ethnic Identity Development and Urban Haitian Adolescents. *Journal of Multicultural Counseling and Development*, *26*(2), 96–109. https://doi.org/10.1002/j.2161-1912.1998.tb00190.x

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social Categorization and the Formation of Intergroup Attitudes in Children. *Child Development*, *68*(3), 530–543. https://doi.org/10.1111/j.1467-8624.1997.tb01956.x

Bigler, R. S., & Liben, L. S. (1993). A Cognitive-Developmental Approach to Racial Stereotyping and Reconstructive Memory in Euro-American Children. *Child Development*, *64*(5), 1507–1518. https://doi.org/10.2307/1131549

Bigler, R. S., & Liben, L. S. (2007). Developmental Intergroup Theory: Explaining and Reducing Children's Social Stereotyping and Prejudice. *Current Directions in Psychological Science*, *16*(3), 162–166. https://doi.org/10.1111/j.1467-8721.2007.00496.x

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Brewer, M. B. (1999). The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *Journal of Social Issues*, *55*(3), 429–444. https://doi.org/10.1111/0022-4537.00126

Brown, C. S., Ali, H., Stone, E. A., & Jewell, J. A. (2017). U.S. Children's Stereotypes and Prejudicial Attitudes toward Arab Muslims: Us Children's Stereotypes of Muslims. *Analyses of Social Issues and Public Policy*, *17*(1), 60–83. https://doi.org/10.1111/asap.12129

Brown, R. (2011). *Prejudice: Its Social Psychology*. John Wiley & Sons.

Calcagno, S., & Dube, W. V. (1994). Emergence of conditional discriminations after constructed-response matching-to-sample training. *Psychological Record*, *44*(4), 509.

Cameron, J. A., Alvarez, J. M., Ruble, D. N., & Fuligni, A. J. (2001). Children's Lay Theories About Ingroups and Outgroups: Reconceptualizing Research on Prejudice. *Personality*

*and Social Psychology Review*, *5*(2), 118–128. https://doi.org/10.1207/S15327957PSPR0502_3

Cameron, L., & Rutland, A. (2006). Extended Contact through Story Reading in School: Reducing Children's Prejudice toward the Disabled. *Journal of Social Issues*, *62*(3), 469–488. https://doi.org/10.1111/j.1540-4560.2006.00469.x

Carvalho, M. P. de, & Rose, J. C. de. (2014). Understanding Racial Attitudes through the Stimulus Equivalence Paradigm. *The Psychological Record*, *64*(3), 527–536. https://doi.org/10.1007/s40732-014-0049-4

Code, C. F. (1976). Proceedings: Suppression of histamine leucocytosis by metiamide. *The Journal of Physiology*, *254*(1), 31P–32P.

Cuéllar, I., Nyberg, B., Maldonado, R. E., & Roberts, R. E. (1997). Ethnic identity and acculturation in a young adult Mexican-origin population. *Journal of Community Psychology*, *25*(6), 535–549. https://doi.org/10.1002/(SICI)1520-6629(199711)25:6<535::AID-JCOP4>3.0.CO;2-O

Cuhadar, E., & Dayton, B. (2011). The Social Psychology of Identity and Inter-group Conflict: From Theory to Practice. *International Studies Perspectives*, *12*(3), 273–293. https://doi.org/10.1111/j.1528-3585.2011.00433.x

De Rose, J. C., Hidalgo, M., & Vasconcellos, M. (2013). Controlling Relations in Baseline Conditional Discriminations as Determinants of Stimulus Equivalence. *Psychological Record*, *63*(1), 85–98.

Dixon, M. R., Rehfeldt, R. A., Zlomke, K. R., & Robinson, A. (2006). Exploring the Development and Dismantling of Equivalence Classes Involving Terrorist Stimuli. *The Psychological Record*, *56*(1), 83–103. https://doi.org/10.1007/BF03395539

Dube, W. V., Green, G., & Serna, R. W. (1993). Auditory Successive Conditional Discrimination and Auditory Stimulus Equivalence Classes. *Journal of the Experimental Analysis of Behavior*, *59*(1), 103–114. https://doi.org/10.1901/jeab.1993.59-103

Enser, M. (1975). Desaturation of stearic acid by liver and adipose tissue from obese-hyperglycaemic mice (ob/ob). *The Biochemical Journal*, *148*(3), 551–555.

Fields, L., Reeve, K. F., Matneja, P., Varelas, A., Belanich, J., Fitzer, A., & Shamoun, K. (2002). The Formation of a Generalized Categorization Repertoire: Effect of Training with Multiple Domains, Samples, and Comparisons. *Journal of the Experimental Analysis of Behavior*, *78*(3), 291–313. https://doi.org/10.1901/jeab.2002.78-291

Glat, R., Gould, K., Stoddard, L. T., & Sidman, M. (1994). A Note on Transfer of Stimulus Control in the Delayed-Cue Procedure: Facilitation by an Overt Differential Response. *Journal of Applied Behavior Analysis*, *27*(4), 699–704. https://doi.org/10.1901/jaba.1994.27-699

Haydu, V. B., Camargo, J., & Bayer, H. (2015). Effects of preexperimental history on the formation of stimulus equivalence classes: A study with supporters of Brazilian soccer clubs. *Psychology & Neuroscience*, *8*(3), 385–396. https://doi.org/10.1037/h0101276

Hayes, L. J., Thompson, S., & Hayes, S. C. (1989). Stimulus Equivalence and Rule Following. *Journal of the Experimental Analysis of Behavior*, *52*(3), 275–291. https://doi.org/10.1901/jeab.1989.52-275

Hayes, L. J., Tilley, K., & Hayes, S. C. (1988). Extending Equivalence Class Membership to Gustatory Stimuli. *The Psychological Record; Granville, Ohio, Etc.*, *38*(4), 473–482.

Holland, J. (1999). *Hope against history: the course of conflict in Northern Ireland* (1st ed). New York: Henry Holt.

Homma, Y., Zumbo, B. D., Saewyc, E. M., & Wong, S. T. (2014). Psychometric Evaluation of the 6-item Version of the Multigroup Ethnic Identity Measure with East Asian Adolescents in Canada. *Identity (Mahwah, N.J.)*, *14*(1), 1–18. https://doi.org/10.1080/15283488.2013.858227

Inzlicht, M., Tullett, A. M., Legault, L., & Kang, S. K. (2011). Lingering Effects: Stereotype Threat Hurts More than You Think. *Social Issues and Policy Review*, *5*(1), 227–256. https://doi.org/10.1111/j.1751-2409.2011.01031.x

Johnson, C., & Sidman, M. (1993). Conditional Discrimination and Equivalence Relations: Control by Negative Stimuli. *Journal of the Experimental Analysis of Behavior*, *59*(2), 333–347. https://doi.org/10.1901/jeab.1993.59-333

Katz, D., & Braly, K. w. (1935). Racial prejudice and racial stereotypes. *Journal of Abnormal and Social Psychology*, *30*(2), 175–193. https://doi.org/10.1037/h0059800

Lam, V. L., & Seaton, J.-A. (2016). Ingroup/Outgroup Attitudes and Group Evaluations: The Role of Competition in British Classroom Settings [Research article]. https://doi.org/10.1155/2016/8649132

Lattal, K., & Perone, M. (2013). *Handbook of Research Methods in Human Operant Behavior*. Springer Science & Business Media.

Leslie, J. C., Tierney, K. J., Robinson, C. P., Keenan, M., Watt, A., & Barnes, D. (1993). Differences Between Clinically Anxious and Non-Anxious Subjects in a Stimulus Equivalence Training Task Involving Threat Words. *The Psychological Record; Granville, Ohio, Etc.*, *43*(1), 153–161.

Liebkind, K., & McAlister, A. L. (1999). Extended contact through peer modelling to promote tolerance in Finland. *European Journal of Social Psychology*, *29*(5–6), 765–780. https://doi.org/10.1002/(SICI)1099-0992(199908/09)29:5/6<765::AID-EJSP958>3.0.CO;2-J

Lynch, D. C., & Cuvo, A. J. (1995). Stimulus Equivalence Instruction of Fraction-Decimal Relations. *Journal of Applied Behavior Analysis*, *28*(2), 115–126. https://doi.org/10.1901/jaba.1995.28-115

Macrae, C. N., Stangor, C., & Hewstone, M. (1996). *Stereotypes and Stereotyping*. Guilford Press.

McAlister, A. L., Ama, E., Barroso, C., Peters, R. J., & Kelder, S. (2000). Promoting tolerance and moral engagement through peer modeling. *Cultural Diversity and Ethnic Minority Psychology*, *6*(4), 363–373. https://doi.org/10.1037/1099-9809.6.4.363

Minneapolis Public Schools. (2017). *racial_ethnic_school_gradefall2017.pdf* (School Demographics). Retrieved from http://studentaccounting.mpls.k12.mn.us/uploads/racial_ethnic_school_gradefall2017.pdf

Minster, S. T., Jones, M., Elliffe, D., & Muthukumaraswamy, S. D. (2006). Stimulus Equivalence: Testing Sidman's (2000) Theory. *Journal of the Experimental Analysis of Behavior*, *85*(3), 371–391. https://doi.org/10.1901/jeab.2006.15-05

Mizael, T. M., de Almeida, J. H., Silveira, C. C., & de Rose, J. C. (2016). Changing Racial Bias by Transfer of Functions in Equivalence Classes. *The Psychological Record*, *66*(3), 451–462. https://doi.org/10.1007/s40732-016-0185-0

Moxon, P. D., Keenan, M., & Hine, L. (1993). Gender-Role Stereotyping and Stimulus Equivalence. *The Psychological Record; Granville, Ohio, Etc.*, *43*(3), 381–393.

Neisser, U. (1989). *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*. CUP Archive.

Paluck, E. L., & Green, D. P. (2009). Prejudice Reduction: What Works? A Review and Assessment of Research and Practice. *Annual Review of Psychology*, *60*(1), 339–367. https://doi.org/10.1146/annurev.psych.60.110707.163607

Patterson, M. M., & Bigler, R. S. (2006). Preschool Children's Attention to Environmental Messages About Groups: Social Categorization and the Origins of Intergroup Bias. *Child Development*, *77*(4), 847–860. https://doi.org/10.1111/j.1467-8624.2006.00906.x

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Phinney, J. S. (1992). The Multigroup Ethnic Identity Measure: A New Scale for Use with Diverse Groups. *Journal of Adolescent Research*, *7*(2), 156–176. https://doi.org/10.1177/074355489272003

Phinney, J. S. (1996). When we talk about American ethnic groups, what do we mean? *American Psychologist*, *51*(9), 918–927. https://doi.org/10.1037/0003-066X.51.9.918

Plaud, J. J. (1995a). The Formation of Stimulus Equivalences: Fear-Relevant Versus Fear-Irrelevant Classes. *The Psychological Record; Granville, Ohio, Etc.*, *45*(2), 207–222.

Plaud, J. J. (1995b). The formation of stimulus equivalences: Fear-relevant versus fear-irrelevant stimulus classes. *Psychological Record*, *45*(2), 207.

Plaud, J. J., & Eifert, G. H. (1998). *From Behavior Theory to Behavior Therapy*. Allyn and Bacon.

Rothman, K. J. (1990). No Adjustments Are Needed for Multiple Comparisons. *Epidemiology*, *1*(1), 43–46.

Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2007). Gender Development. In *Handbook of Child Psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470147658.chpsy0314

Rutland, A., Nesdale, D., & Brown, C. S. (2017). *The Wiley Handbook of Group Processes in Children and Adolescents*. John Wiley & Sons.

Sidman, M, Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior*, *43*(1), 21–42. https://doi.org/10.1901/jeab.1985.43-21

Sidman, Murray. (1990). Equivalence relations: Where do they come from? In *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 93–114). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Sidman, Murray. (1994). *Equivalence relations and behavior: A research story* (Vol. xvii). Boston, MA, US: Authors Cooperative.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797.

Stikeleather, G., & Sidman, M. (1990). An instance of spurious equivalence relations. *The Analysis of Verbal Behavior*, *8*, 1–11.

Tierney, K. J., & De Largy, P. (1995). Formation of an equivalence class incorporating haptic stimuli. *Psychological Record*, *45*(3), 431.

Tierney, K. J., Largy, P. D., & Bracken, M. (1995). Formation of an Equivalence Class Incorporating Haptic Stimuli. *The Psychological Record*, *45*(3), 431–437. https://doi.org/10.1007/BF03395152

Toussaint, K. A., & Tiger, J. H. (2010). Teaching early braille literacy skills within a stimulus equivalence paradigm to children with degenerative visual impairments. *Journal of Applied Behavior Analysis*, *43*(2), 181–194. https://doi.org/10.1901/jaba.2010.43-181

Tropp, L. R., & Bianchi, R. A. (2007). Interpreting references to group membership in context: feelings about intergroup contact depending on who says what to whom. *European Journal of Social Psychology*, *37*(1), 153–170. https://doi.org/10.1002/ejsp.340

Vezzali, L., Stathi, S., & Giovannini, D. (2012). Indirect contact through book reading: Improving adolescents' attitudes and behavioral intentions toward immigrants. *Psychology in the Schools*, *49*(2), 148–162. https://doi.org/10.1002/pits.20621

Watt, A., Keenan, M., Barnes, D., & Cairns, E. (1991). Social Categorization and Stimulus Equivalence. *The Psychological Record; Granville, Ohio, Etc.*, *41*(1), 33–50.

Williams, J. E., Best, D. L., Boswell, D. A., Mattson, L. A., & Graves, D. J. (1975). Preschool racial attitude measure II. *Educational and Psychological Measurement*, *35*(1), 3–18.

Zentall, T. R., & Smeets, P. M. (1996). *Stimulus Class Formation in Humans and Animals*. Elsevier.

Appendix A

**MEIM**

6 Item, adapted MEIM (multi-group ethnic identify measure)

In this country, people come from many different countries and cultures, and there are many different words to describe the different backgrounds or ethnic groups that people come from. Some examples of the names of ethnic groups are American Indian, Native American, Dakota, Lakota, Hispanic or Latino, East African, Somali, Somali American, Ethiopian, Ethiopian American, Asian American, Black or African, American, Chinese, Filipino, Mexican American, Caucasian or White, Italian American, and many others. These questions are about your ethnicity or your ethnic group and how you feel about it or react to it.

Please fill in: In terms of ethnic group, I consider myself to be _____

Use the numbers below to indicate how much you agree or disagree with each statement.

**(4) Strongly agree (3) Agree (2) Disagree (1) Strongly disagree**

1- I have spent time trying to find out more about my ethnic group, such as its history, traditions, and customs.
2- I participated in cultural practices of my own group, such as special good, music, or customs.

3- I have often talked to other people in order to learn more about my ethnic group.
4- I have a strong sense of belonging to my ethnic group.
5- I understand what my ethnic group membership means to me.
6- I feel a strong attachment towards my own ethnic group.

Appendix B

**TRAIT RATINGS MEASURE**

For each selection: the following four choices are available:
4 (all of the students at _____)
3 (many of the students at _____)
2 (some of the students at _____)
1 (none of the students at _____)

1.Picture the students at Wiconi. How many of the students at Wiconi would you say are friendly?
2.Picture the student at Canug. How many of the students at Canug would you say are good?
3.Picture the students at Wiconi. How many of the students at Wiconi would you say are dirty?
4.Picture the student at Canug. How many of the students at Canug would you say are mean?
5.Picture the students at Wiconi. How many of the students at Wiconi would you say are nice?
6.Picture the student at Canug. How many of the students at Canug would you say are helpful?
7.Picture the students at Wiconi. How many of the students at Wiconi would you say are naughty?
8.Picture the student at Canug. How many of the students at Canug would you say are selfish?
9.Picture the students at Wiconi. How many of the students at Wiconi would you say are smart?
10.Picture the student at Canug. How many of the students at Canug would you say are unfriendly?
11.Picture the students at Wiconi. How many of the students at Wiconi would you say are good?
12.Picture the student at Canug. How many of the students at Canug would you say are friendly?
13.Picture the students at Wiconi. How many of the students at Wiconi would you say are mean?
14.Picture the student at Canug. How many of the students at Canug would you say are dirty?
15.Picture the students at Wiconi. How many of the students at Wiconi would you say are helpful?
16.Picture the student at Canug. How many of the students at Canug would you say are nice?
17.Picture the students at Wiconi. How many of the students at Wiconi would you say are selfish?

18.Picture the student at Canug. How many of the students at Canug would you say are naughty?
19.Picture the students at Wiconi. How many of the students at Wiconi would you say are unfriendly?
20.Picture the student at Canug. How many of the students at Canug would you say are smart?

Appendix C

**INTERGROUP ATTITUDE MEASURE**

***Intergroup Attitude Measure.*** Each item will be presented in a 6 point response scale that consists of:

1-Very Untrue of what I believe
2-Untrue of what I believe
3-Somewhat untrue of what I believe
4-Somewhat true of what I believe
5-True of what I believe
6-Very true of what I believe

For each of these items, the word "outgroup" will be replaced in the actual question with either "Anne Sullivan students" or "Anishinabe Students" based on the school program of the subject being questioned. These items are:

1. "I would like it if there were more OUTGROUP MEMBERS here (in the building)."

2. "I do not want OUTGROUP MEMBERS in my classroom" (reverse-scored).

3. "OUTGROUP MEMBERS cause a lot of problems like fights and bullying in this building." (reverse-scored)

4. "OUTGROUP MEMBERS strongly increase problems like crime, drugs, illnesses etc. (reverse-scored)

5. "I would not mind to have an OUTGRUOP MEMBER as my teacher."

6. "I do not approve of using names that might hurt OUTGROUP MEMBERS."

7. "IF there are problems like fights and bullying in this building than students like OUTGROUP MEMBERS should not be allowed to come here." (reverse-scored)

8. "I think it is very natural that people are friends with OUTGROUP MEMBERS"

9. “OUTGROUP MEMBERS should be able to follow their own customs without being bullied or teased.”

10. “I think it would be great to have more OUTGROUP MEMBERS living near where I live.”

11. “If I get mad, I may call OUTGROUP MEMBERS bad names referring to their skin color or where their family comes from.” (reverse-scored)

Appendix D

## INGROUP AND OUTGROUP BEHAVIORAL INTENTIONS

Students will respond to the questions using one of the following four choices: Very happy (4), happy (3), unhappy (2), or very unhappy (1).

Scenario #1: A new student has just moved here and it is their first day of school. You have never met this person before. The student is then shown an image of a Native American student (as selected by the Native American stimuli pilot) and asked:

Question #1: How happy will you be to meet him or her?
Question #2: How happy would you be to play with him or her?
Question #3: How happy would you be to sit next to them at the lunch table?
Question #4: How happy would you be to work on a class assignment or do schoolwork with him or her?

Scenario #2: A new student has just moved here and it is their first day of school. You have never met this person before. The student is then shown an image of an East African student (as selected by the East African stimuli pilot) and asked:

Question #1: How happy will you be to meet him or her?
Question #2: How happy would you be to play with him or her?
Question #3: How happy would you be to sit next to them at the lunch table?
Question #4: How happy would you be to work on a class assignment or do schoolwork with him or her?

Appendix E

## PARENT CONSENT FORM

## PARENT CONSENT FORM

The Effects of Stimulus Equivalence Training on Intercultural Relations

Your child is invited to be part of a project to help us learn how to promote positive interactions between students from different cultural backgrounds. Your child was selected as a possible participant because he or she is a student in a Minneapolis Public School. We ask that you read this form and ask any questions you may have before allowing your child to participate in the project.

Dr. Jennifer J. McComas and graduate students from the University of Minnesota are conducting this project.

**Background Information:**

The purpose of this project is to develop knowledge about how the use of match-to-sample procedures (teaching by having kids match pictures) might be used to promote positive and helpful interactions between students from different cultural backgrounds.

**Procedures:**

If you allow your child to be part of this project, he or she will be asked to look at some pictures of other students and rate the pictures. They will also be asked to match pictures of students with other symbols. There are no right or wrong answers, we are simply looking to determine how students match different kinds of pictures. If they are selected to take part in the training for the project, they will be asked to go through a 3-5 sessions of matching pictures to other pictures. Your child may earn small rewards for taking part in successfully matching the pictures.

Data collected during this project will not be able to be identified to your child. We will not share any information about any child with anyone.

**Risks and Benefits of being in the Project**:

We do not anticipate any risk to any student who participates in the project.

The benefits to participation are:

There are not certain benefits to your child for participating in the project. It is possible that children that participate will be more likely to talk with and play with new and different students in the after school program.

Compensation: Your child will receive no payment for his or her participation. However, all students in the project may receive stickers, bookmarks, snacks, or other small items to help them keep track of their progress or reward them when they accomplish their goals.

**Confidentiality:**

During the project there is a small risk of loss of privacy. To address this risk the records and data from this project will be kept private and all project information will be stored securely in a locked room or cabinet at the University of Minnesota, where only individuals associated with the project will have access to the information.

You child will also be given a project number. Your information will be linked to this number and stored in a secure, separate location at the University of Minnesota and only project staff will have access to the records.
There will be no identifying information used in any publications of this project. In any sort of report that we might publish, we will not include any information that will make it possible to identify your child.
Paperwork used in the project (such as forms or observation sheets) will be shredded.

**Voluntary Nature of the Project:**
Your participation in this project is voluntary. Your decision whether or not to allow your child to participate will not affect your current for future relations with the University of Minnesota or Minneapolis Public Schools. The alternative to participating in this project is to decline participation. You can say, "NO" to participation in this project. If you decide to say "YES" and participate, you are free to withdraw at any time without affecting your current or future relations with the University of Minnesota or Minneapolis Public Schools. Your child will still be able to attend his or her current school and after school program even if you do not allow him or her to participate in the project.
If you check "YES" and allow your child to participate in the project, we will be able to use your child's data to better understand how to create positive and accepting environments for children in schools. In addition we will collect demographic information about your child including his or her age, grade, home language, and race/ethnicity. This information will be used to describe the range of children who participate in the project and we will NOT share any information about individual children, only as a whole group of participants.
If you check "NO" your child will still receive regular after classroom instruction and after school care.
Please check either "YES" or "NO", write your child's name on the form, sign it, and send it back with your child.
Robert Henery from the University of Minnesota is available to answer any questions you may have.

**Contacts and Questions:**
Contacts and questions can be directed to Robert Henery at the University of Minnesota. You may ask any questions at anytime. If you have questions later, you are encouraged to contact him at 250 Education Sciences 56 East River Road, Minneapolis, MN 55455; 612-624-0037; hener005@umn.edu
If you have any questions or concerns regarding this project and would like to talk to someone other than the project staff, you are encouraged to contact the Research

Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast Minneapolis, MN 55455; (612)-625-1650.

You will be given a copy of this information to keep for your records.

**Statement of Consent:**

I have read the above information. I have asked questions and have received answers.

☐ YES I agree to allow my child to participate in the project.

☐ NO I do not want my child to participate in the project

_____

Name of student

_____

Parent Signature

Appendix F

**PARENT CONSENT FORM TRANSLATED INTO SOMALI**

**FOOMKA OGOLAASHADA WAALIDIINTA**

Saameynta Tababarka Isbeddelka Tababbarka ee Xiriirrada Deegaanka

Ilmahaagu waxa lagu martiqaaday inay qayb ka noqdaan mashruuc si ay nooga caawiyaan inaan barto sida loo kobcin lahaa isdhexgalka wanaagsan ee ardayda ka soo jeeda dhaqamada kala duwan. Ilmahaaga ayaa loo doortay inuu yahay kaqeybgal suurtagal ah maxaa yeelay isaga ama iyadu waa arday dhigta Dugsiyada Dadweynaha Minneapolis. Waxaan kaa codsanaynaa in aad akhrisid foomkan oo aad waydiiso wixii su'aalo ah ee aad qabi kartid ka hor intaadan u ogolaanin ilmahaaga inuu ka qaybgalo mashruuca.

**Dr. Jennifer J. McComas** iyo ardayda ka qalin-jebisa Jaamacadda Minnesota ayaa sameeya mashruucan.

## Macluumaadka Hore:

Ujeedada mashruucan ayaa ah in la horumariyo aqoonta ku saabsan sida loo isticmaalo hababka loo isticmaalo habka loo yaqaan 'sample-to-sample' (barashada sawir qaadashada sawirro) ayaa loo isticmaali karaa in lagu dhiirigeliyo isdhexgalka wanaagsan ee waxtarka leh ee ardayda ka soo jeeda dhaqamada kala duwan.

## Nidaamka:

Haddii aad u ogolaato inuu ilmahaagu qayb ka noqdo mashruucan, isaga ama iyada waxaa laga codsan doonaa in ay eegaan sawirro qaar ka mid ah ardayda kale oo qiimeynaya sawirada. Waxaa sidoo kale lagu weydiin doonaa inay sawirro muuqaalo ardayda leh calaamado kale. Ma jiraan jawaabo sax ah ama khalad ah, waxaanu si fudud u eegaynaa sida loo ogaado sida ardaydu u

habeeyaan noocyo kala duwan oo sawiro ah. Haddii loo doorto inay kaqeyb qaataan tababbarka mashruuca, waxaa la weydiin doonaa in ay maraan 3-5 kulan oo sawiro u dhigma sawirada kale. Ilmahaagu wuxuu ka heli karaa abaalmarin yar si uu uga qayb qaato si guul leh u-habboon sawirada.

Waxaanu u eegi doonaa ardayda iyo sida ay ula dhaqmayaan midba midka kale dabiiciga barnaamijka dabiiciga ah ee dugsiga kadib. Xogta la ururiyey inta lagu jiro kormeerka lama ogaanayo ilmahaaga. Xogtaasi waxay ku kooban tahay waxa ilmahaagu sameeyo waqtiga cuntada fudud ee barnaamijka Beacons. Ma wadaagi doonno macluumaad kasta oo ku saabsan ilmo kasta oo aan cidna la wadaagin.

## Khatarta iyo Faa'iidooyinka Mashruuca:

Ma filayno khatar kasta arday kasta oo ka qaybgalaya mashruuca.

Faa'iidooyinka ka-qaybgalka:

Ma jiraan wax faa'iido u leh ilamahaaga si aad uga qayb qaadato mashruuca. Waxaa suurtagal ah in carruurta ka qeybgaleysa ay u badan tahay inay la hadlaan oo ay la ciyaaraan ardayda cusub iyo kuwa kala duwan ee barnaamijka dugsi kadib.

Magdhawga: Ilmahaagu ma heli doono wax lacag bixin ah ee kaqeybqaadashadiisa. Si kastaba ha noqotee, dhamaan ardayda mashruuca waxaa laga yaabaa inay helaan warqado, warqado, cunto fudud, ama waxyaabo kale oo yaryar si ay uga caawiyaan inay sii wataan horumarkooda ama abaal mariyaan marka ay fuliyaan ujeedooyinkooda.

## Sirta:

Inta lagu guda jiro mashruuca waxaa jira halis yar oo ah khasaare khaas ah. Si wax looga qabto khatarta diiwaanada iyo xogta mashruucan waxaa loo hayn doonaa kuwo gaar ah oo dhammaan macluumaadka mashruuca waxaa lagu keydin doonaa si qarsoodi ah qolka qufulan ama golaha ee

Jaamacadda Minnesota, halkaas oo shakhsiyaadka la xidhiidha mashruuca ay heli doonaan macluumaadka.

Ilmahaaga waxaa lagu siin doonaa lambar mashruuc. Macluumaadkaagu wuxuu ku xiran yahay lambarkan oo lagu keydiyaa meel amaan ah oo ka baxsan Jaamacadda Minnesota iyo shaqaalaha kaliya ee mashruuca ayaa heli doona diiwaanada.

Ma jiri doono macluumaad aqoonsi oo loo adeegsado daabacaad kasta ee mashruucan. Warbixin kasta oo laga yaabo in aan daabacno, kuma koobnaan doono macluumaad kasta oo suurta galinaya in la aqoonsado ilmahaaga.

Qorshaha waraaqaha ee loo isticmaalo mashruuca (sida foomamka ama xaashida kormeerka) waa la burburin doonaa

## Nooca iskaa wax u qabso ee mashruuca:

Ka qaybqaadashadaada mashruucan waa ikhtiyaari. Go'aankaaga haddii ay tahay ama u ogolaan lahayd ilmahaagu inuu ka qaybqaato ma saameyn doonto xaadirkaaga mustaqbalka ee xiriirka mustaqbalka ee Jaamacadda Minnesota ama Dugsiyada Dadweynaha Minneapolis. Beddelka ka-qaybgalka mashruucan waa in la yareeyo ka-qaybgalka. Waxaad dhihi kartaa, "MAYA" inaad ka qayb qaadato mashruucan. Haddii aad go'aansato inaad "HAA" ka dhigato oo aad ka qayb qaadato, waxaad xor u tahay inaad ka baxdo wakhti kasta iyada oo aan saameyn ku yeelanaynin xiriirkaaga hadda ama mustaqbalkaaga ee Jaamacadda Minnesota ama Dugsiyada Dadweynaha Minneapolis. Ilmahaagu weli wuu awoodi karaa inuu dhigto iskuulkiisa ama barnaamijkiisa dugsi kadib xitaa haddii aadan u ogolaanin isaga ama iyada inay ka qaybgalaan mashruuca. Haddii aad hubiso "HAA" oo u oggolow ilmahaaga inuu ka qaybgalo mashruuca, waxaan awoodi doonaa inaanu isticmaalno xogta ilmahaaga si aan u fahanno sida loo abuuro jawi wanaagsan oo aqbalaya jawi carruurta ee dugsiyada. Intaas waxaa dheer oo aan soo qaadeynaa macluumaadka dadweynaha ee ku saabsan ilmahaaga oo ay ku jiraan da'diisa, fasalka, luqadda guriga, iyo

jinsiyada / jinsiga. Macluumaadkan waxaa loo isticmaali doonaa in lagu qeexo tirada carruurta ka qaybqaata mashruucan, isla markaasna ma wadaageyno wax macluumaad ah oo ku saabsan carruurta shakhsiyadeed, oo keliya koox ahaan ka qaybgalayaasha. Haddii aad calaamadeyso "MAYA" ilmahaagu wuxuu helayaa wakhti caadi ah oo casharro fasal ah iyo daryeel dugsi kadib. Fadlan calaamadee "HAA" ama "MAYA", ku qor magaca ilmahaaga foomka, saxiix, kuna soo celi ilmahaaga. Robert Henery oo ka socda Jaamacadda Minnesota ayaa diyaar u ah inuu ka jawaabo su'aal kasta oo aad qabto.

## Xiriirrada iyo Su'aalaha:

Xiriirrada iyo su'aalaha waxaa lagu hagaajin karaa Robert Henery ee Jaamacadda Minnesota. Waad weydiisan kartaa wixii su'aalo ah wakhti kasta. Haddii aad su'aal qabtid ka dib, waxaa lagugu dhiirigelinayaa inaad la xiriirto 250 Waxbarashada Waxbarashada **56 East River Road, Minneapolis, MN 55455; 612-624-0037; hener005@umn.edu** Haddii aad hayso wax su'aalo ah ama walaac ah oo ku saabsan mashruucaan oo aad jeclaan lahayd inaad la hadasho qof aan ka ahayn shaqaalaha mashruuca, waxaa lagugu dhiirigelinayaa inaad la xiriirto Qeybta Cilmi-baarista Mawduuca, D528 Mayo, 420 Delaware St. Southeast Minneapolis, MN 55455; (**612) -625-1650**. Waxaa lagu siin doonaa nuqulka xogtan si aad u sii haysato diiwaannadaada. **Qoraalka**

## Oggolaanshaha:

Waan aqriyay macluumaadka kor ku xusan. Waxaan su'aalo waydiiyay oo aan helay jawaabaha

☐ **HAA** waxaan oggolahay inaan ogolaado in ilmahaygu ka qaybgalo mashruuca.

☐ **Maya** ma rabo in ilmahaygu ka qaybqaato mashruuca

Student Name

Parent Signature

Appendix G

**STUDENT ASSENT FORM**

We are asking you to help us learn about how to help students be more friendly with each other. Because you are a student in a Minneapolis Public School, we are asking if you want to be in a project.

If you agree to be in this project, we will use your ratings and choices on matching pictures. We hope that this information will help us learn how to get different people to be nice to each other.

You will have the same classes and the same schedule if you say no or yes to being in the project. No program staff will know whether you have said yes or no. If you decide to do the project and later change your mind, you can stop participating at any time. Being in this project is up to you and no one will be mad if you don't want to do it.

Signing your name means you have read this paper or had it read to your, and are willing to be in the project. If you don't want to be in the project, sign by the NO line. Remember, being in this project is your choice and you can change your mind later.

☐ YES I agree to participate in the project

☐ NO I do not want to participate in the project.

Signature of participant _____

Signature of person explaining this project:

_____

Date: _____

Appendix H

**STUDENT ASSENT FORM TRANSLATED INTO SOMALI**

Warqada ogolaanshaha ardayga

Waxaan ku waydiisaneynaa inaad naga caawiso in aan barano sida loo-caawiyo ardayda sida ay u noqdaan dad saxiibtinimo is-tusa. Maadaba aad tahay arday dhigta Dugsiyada Dadweynaha ee Minneapolis, waxaan ku waydiisanaynaa inaad rabtid inaad ka qeybqaadatid mashruuc. Haddii aad ogolaato inaad ka qeybqaadato mashruucan, waxaan isticmaali doonaa xulashadada iyo qiyaastaada ee ku-saabsan isku aadinta sawirada. Waxaan rajeyneynaa in maclumaadkan naga caawin doonto inaan barano sida dadka kala du-duwan isku ixtiraamaan. Waxaad qaadan doontaa fasalo isku mid ah haddii aad diido iyo hadii aad yeesho inaad ka qeybqaadatid mashruuca. Ma jiro shaqaale ogaan doona inaad yeeshay iyo inaad diiday. Haddii aad go'aansatay inaan ka qeybqaadato mashruuca lakiin aad maskaxdaada badashay, waad joojin kartaa ka-qeybqaadashada waqti walba. Ka qeybqaadashada mashruucan waa mid adiga kaa filanta mana jiro qof kaaga xanaaqi doona haddii aadan rabin inaad sameeyso. Magacaaga saxiixidiisa ayaa ah inaad akhrisay warqadan ama laguu aqriyay, diyaarna aad u tahay inaad sameyso. Haddii aadan rabin inaad ka qeybqaadato mashruucan , saxiix meesha Maya ku qoran. Xasuusnow, in ka qeybqaadashada mashruucan uu yahay mid adiga go'aan u leh aadna gadaal maskaxdaada ka bedeli kartid.

☐ Haa Waan oggolahay inaan ka qeybqaato mashruuca.

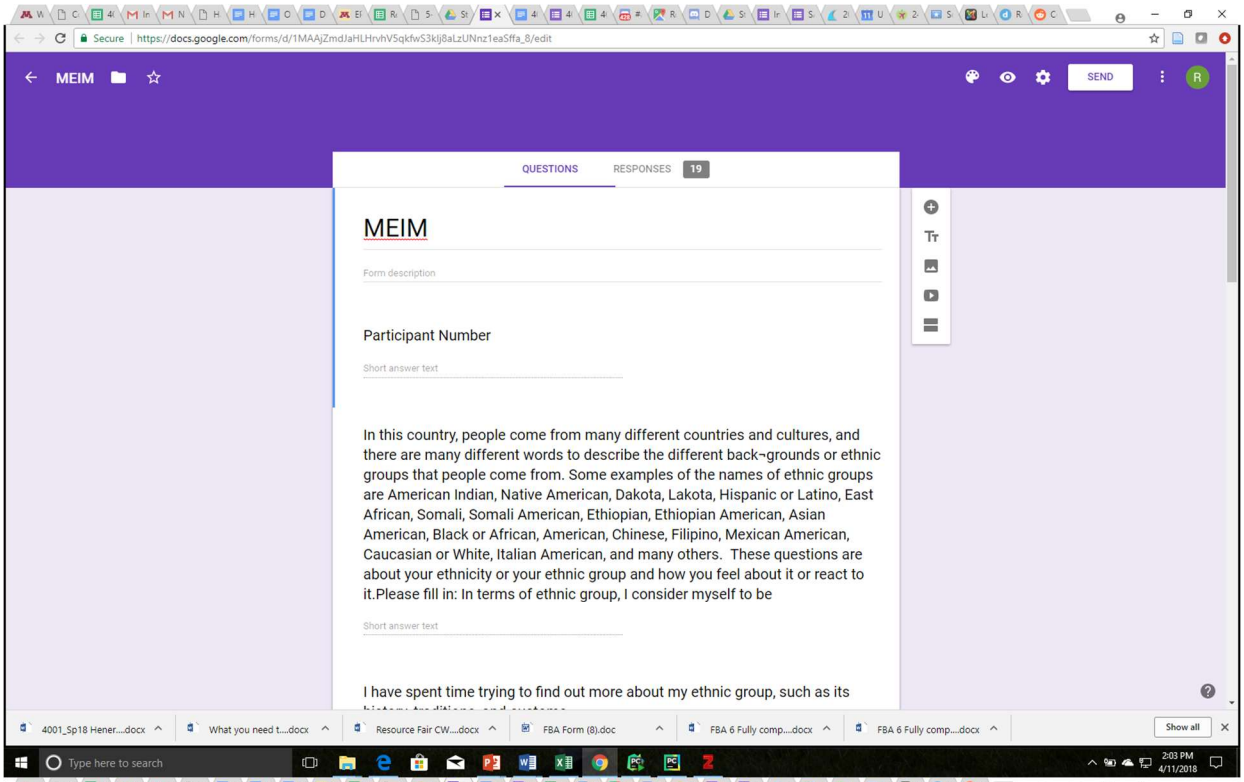
☐ Maya Ma rabo inaan ka qeybqaato mashruuca.

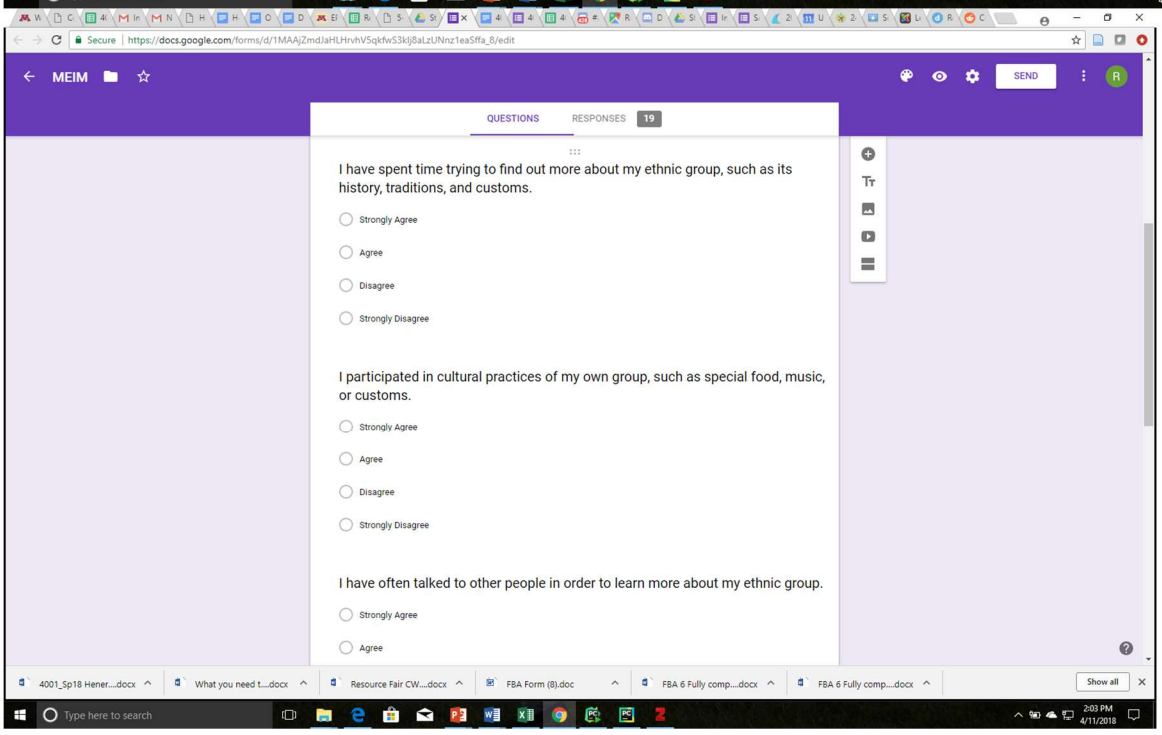
Saxiixa Kaqeybgalaha: _____

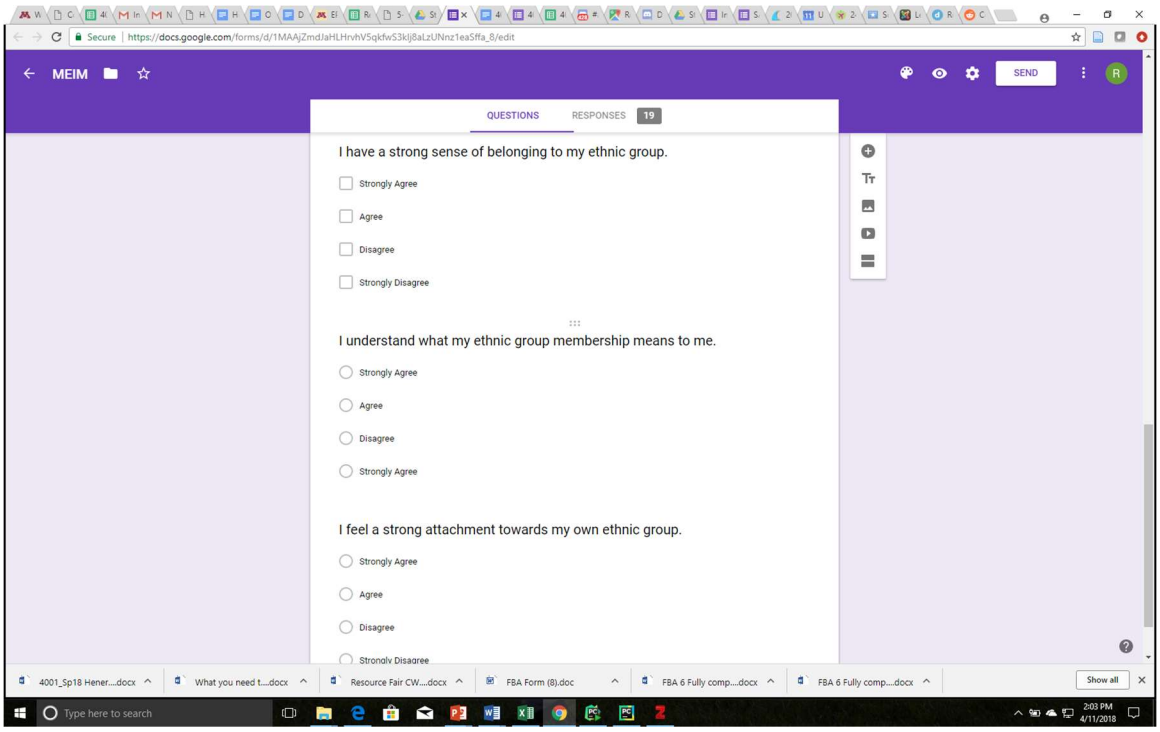
Saxiixa qofka sharaxaya mashruucan: _____

Taariikhda: _____

# Appendix I
# MEIM MEASURE AS SEEN BY PARTICIPANTS

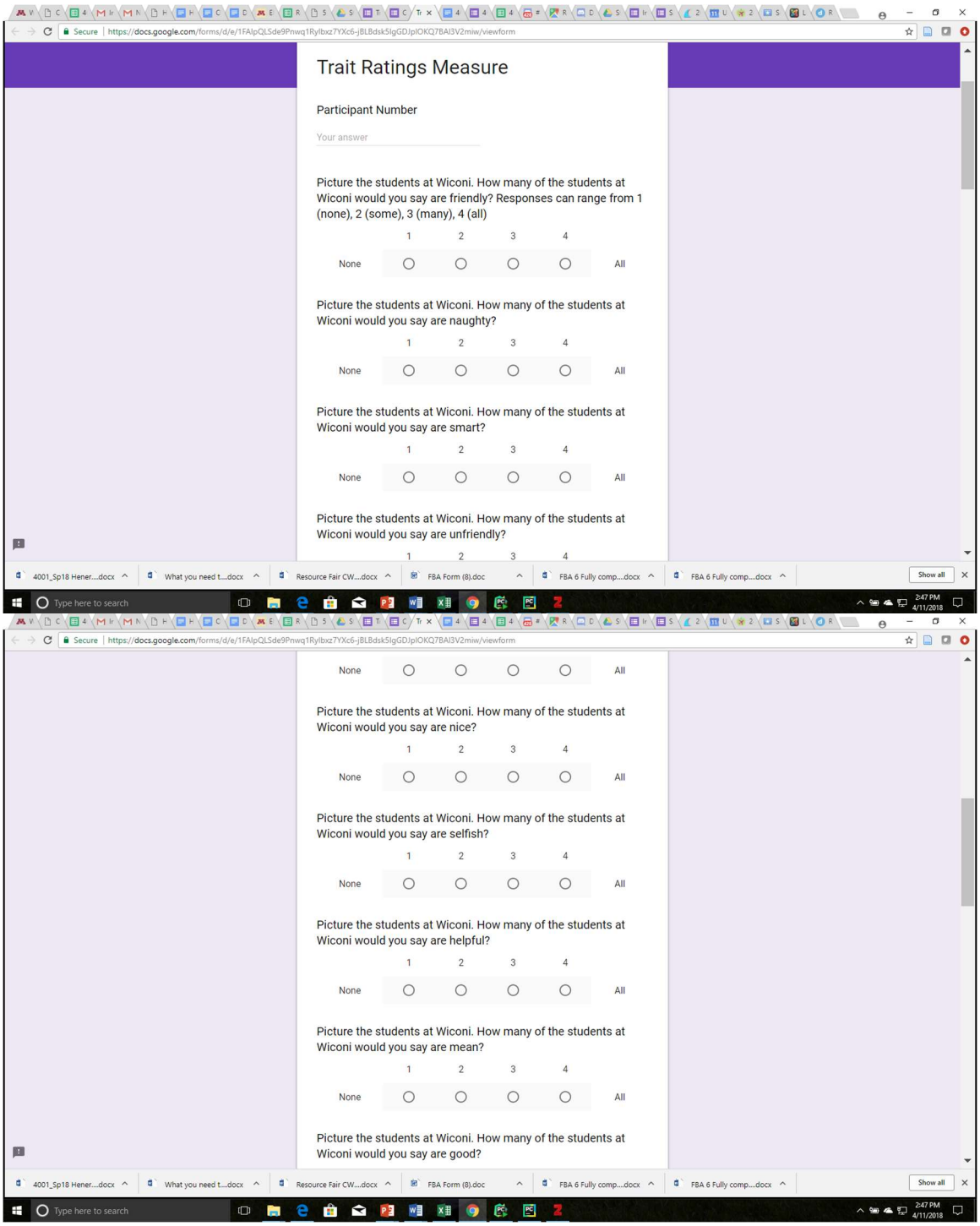


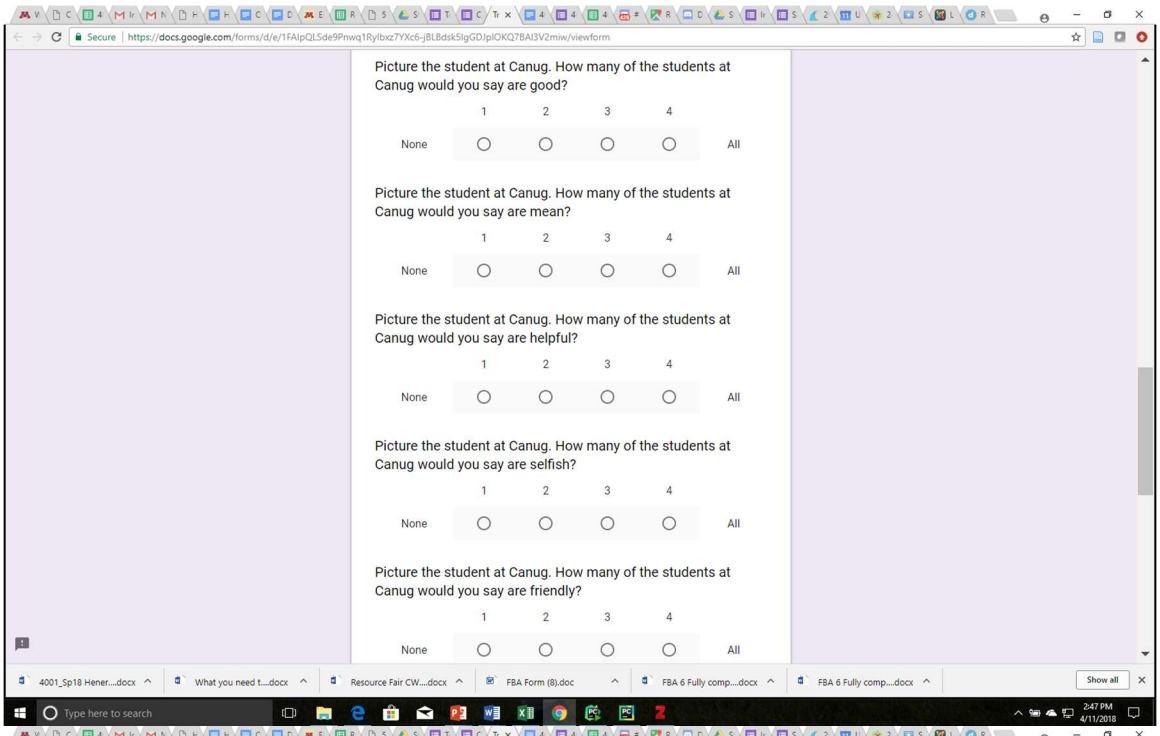




Secure https://docs.google.com/forms/d/1MAAjZmdJaHLHrvhV5qkfwS3klj8aLzUNnz1eaSffa_8/edit
MEIM
SEND
QUESTIONS
RESPONSES 19
I have a strong sense of belonging to my ethnic group.
Strongly Agree
Agree
Disagree
Strongly Disagree
I understand what my ethnic group membership means to me.
Strongly Agree
Agree
Disagree
Strongly Agree
I feel a strong attachment towards my own ethnic group.
Strongly Agree
Agree
Disagree
Strongly Disagree
4001_Sp18 Hener....docx
What you need t....docx
Resource Fair CW....docx
FBA Form (8).doc
FBA 6 Fully comp....docx
FBA 6 Fully comp....docx
Show all
Type here to search
2:03 PM
4/11/2018

Appendix J

**TRAIT RATINGS MEASURE AS SEEN BY PARTICIPANTS**





Picture the student at Canug. How many of the students at Canug would you say are good?

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| None | ○ | ○ | ○ | ○ | All |

Picture the student at Canug. How many of the students at Canug would you say are mean?

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| None | ○ | ○ | ○ | ○ | All |

Picture the student at Canug. How many of the students at Canug would you say are helpful?

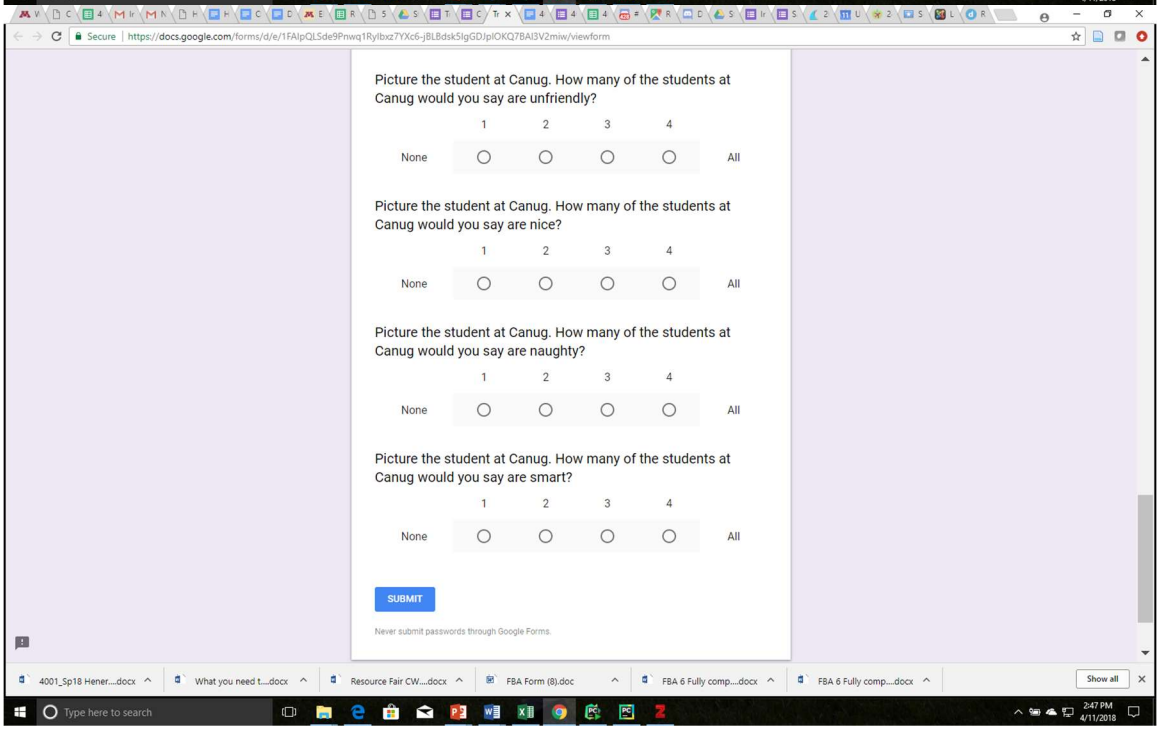
| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| None | ○ | ○ | ○ | ○ | All |

Picture the student at Canug. How many of the students at Canug would you say are selfish?

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| None | ○ | ○ | ○ | ○ | All |

Picture the student at Canug. How many of the students at Canug would you say are friendly?

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| None | ○ | ○ | ○ | ○ | All |



Picture the student at Canug. How many of the students at Canug would you say are unfriendly?

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| None | ○ | ○ | ○ | ○ | All |

Picture the student at Canug. How many of the students at Canug would you say are nice?

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| None | ○ | ○ | ○ | ○ | All |

Picture the student at Canug. How many of the students at Canug would you say are naughty?

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| None | ○ | ○ | ○ | ○ | All |

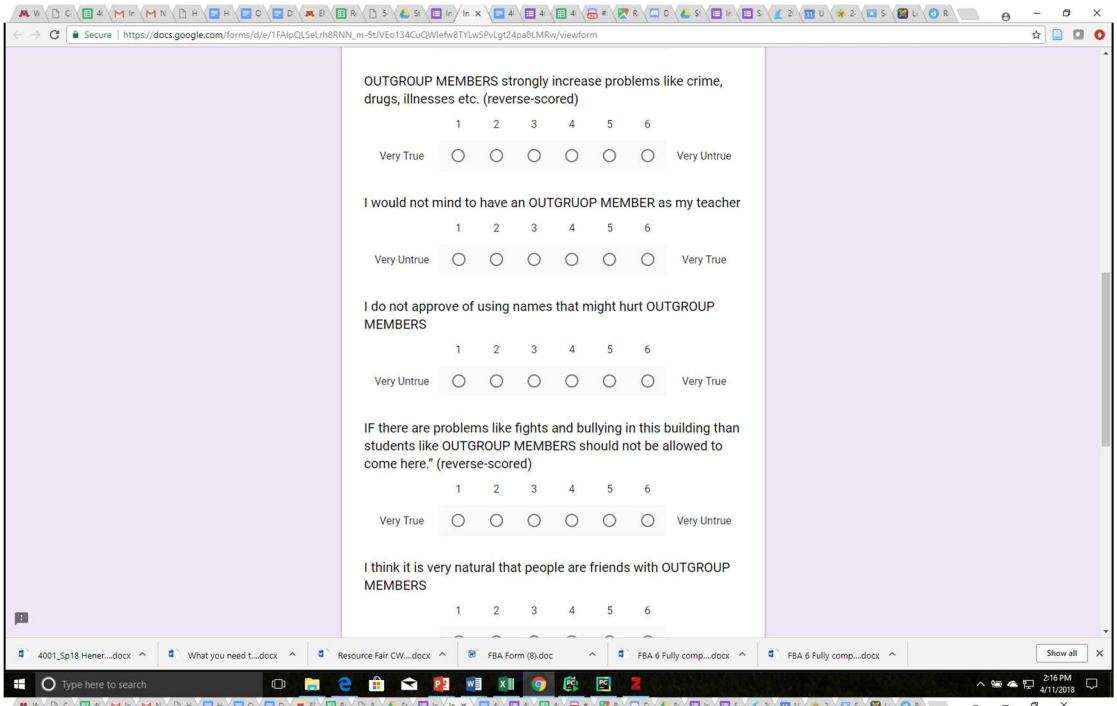
Picture the student at Canug. How many of the students at Canug would you say are smart?

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| None | ○ | ○ | ○ | ○ | All |

SUBMIT

Never submit passwords through Google Forms.

# Appendix K
# INTERGROUP ATTITUDES MEASURE AS SEEN BY PARTICIPANTS



OUTGROUP MEMBERS strongly increase problems like crime, drugs, illnesses etc. (reverse-scored)

| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| Very True | ○ | ○ | ○ | ○ | ○ | ○ | Very Untrue |

I would not mind to have an OUTGRUOP MEMBER as my teacher

| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| Very Untrue | ○ | ○ | ○ | ○ | ○ | ○ | Very True |

I do not approve of using names that might hurt OUTGROUP MEMBERS

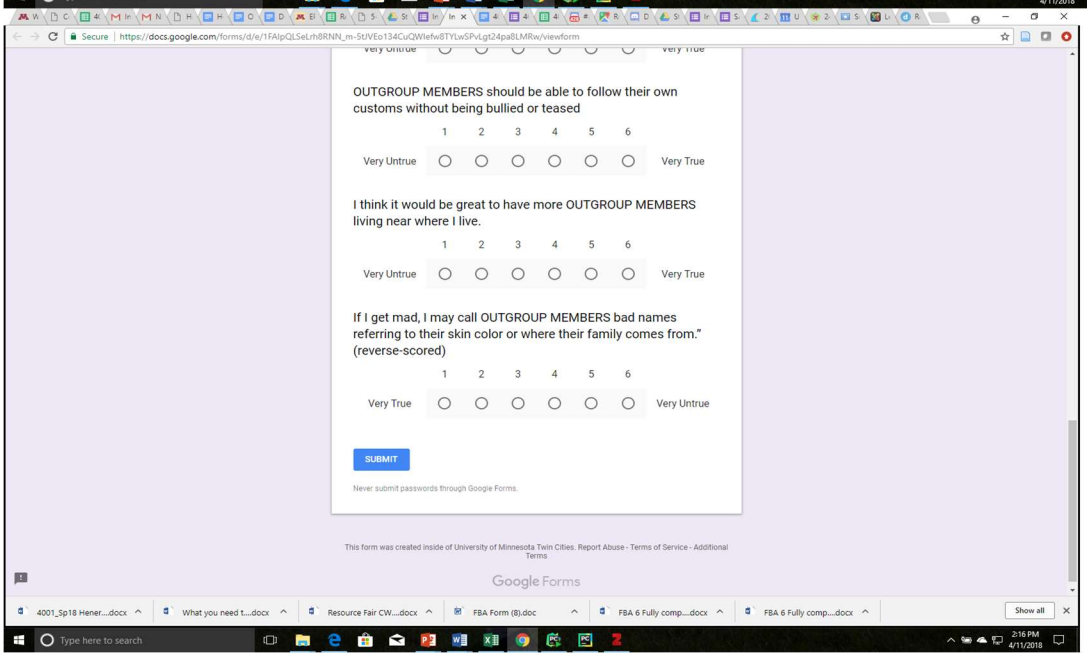
| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| Very Untrue | ○ | ○ | ○ | ○ | ○ | ○ | Very True |

IF there are problems like fights and bullying in this building than students like OUTGROUP MEMBERS should not be allowed to come here." (reverse-scored)

| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| Very True | ○ | ○ | ○ | ○ | ○ | ○ | Very Untrue |

I think it is very natural that people are friends with OUTGROUP MEMBERS

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|



OUTGROUP MEMBERS should be able to follow their own customs without being bullied or teased

| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| Very Untrue | ○ | ○ | ○ | ○ | ○ | ○ | Very True |

I think it would be great to have more OUTGROUP MEMBERS living near where I live.

| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| Very Untrue | ○ | ○ | ○ | ○ | ○ | ○ | Very True |

If I get mad, I may call OUTGROUP MEMBERS bad names referring to their skin color or where their family comes from." (reverse-scored)

| | 1 | 2 | 3 | 4 | 5 | 6 | |
|---|---|---|---|---|---|---|---|
| Very True | ○ | ○ | ○ | ○ | ○ | ○ | Very Untrue |

SUBMIT

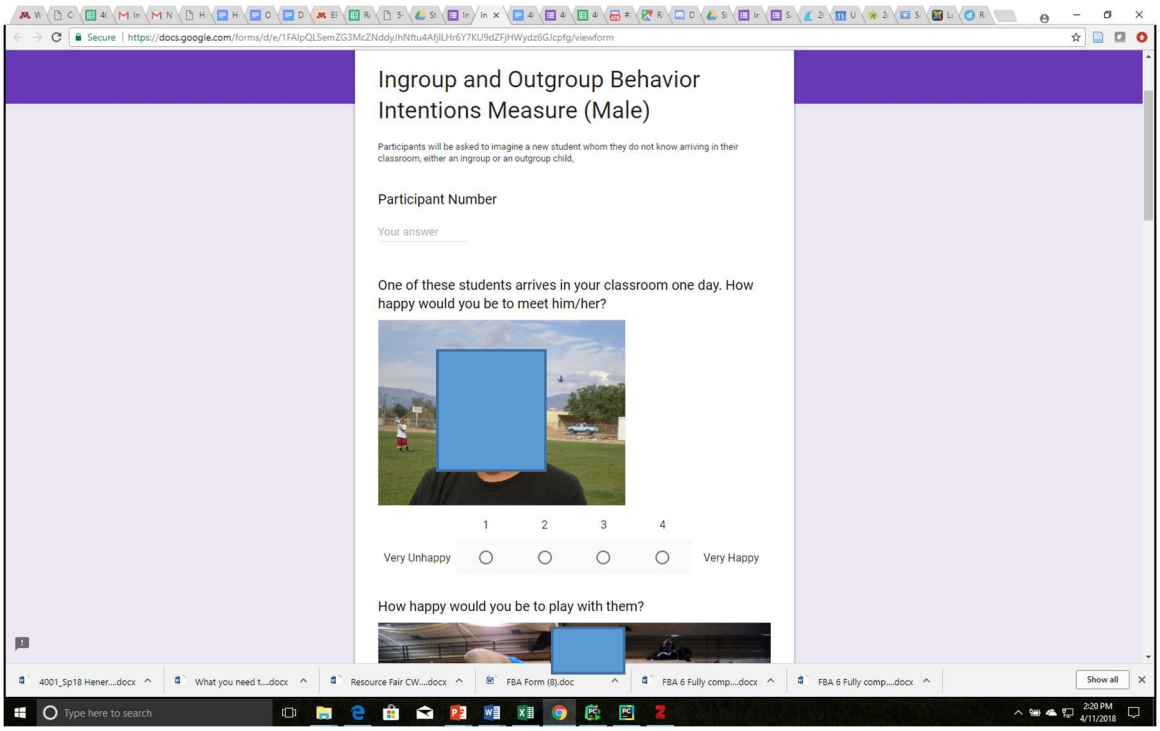
Never submit passwords through Google Forms.

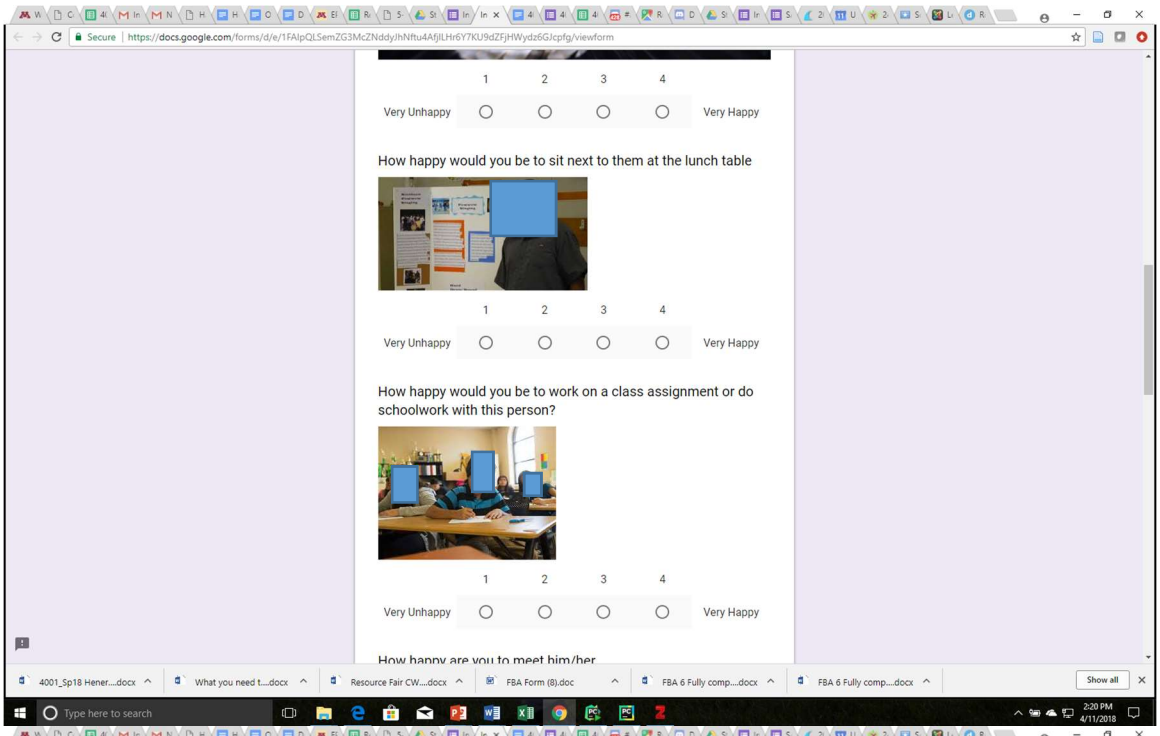
This form was created inside of University of Minnesota Twin Cities. Report Abuse - Terms of Service - Additional Terms

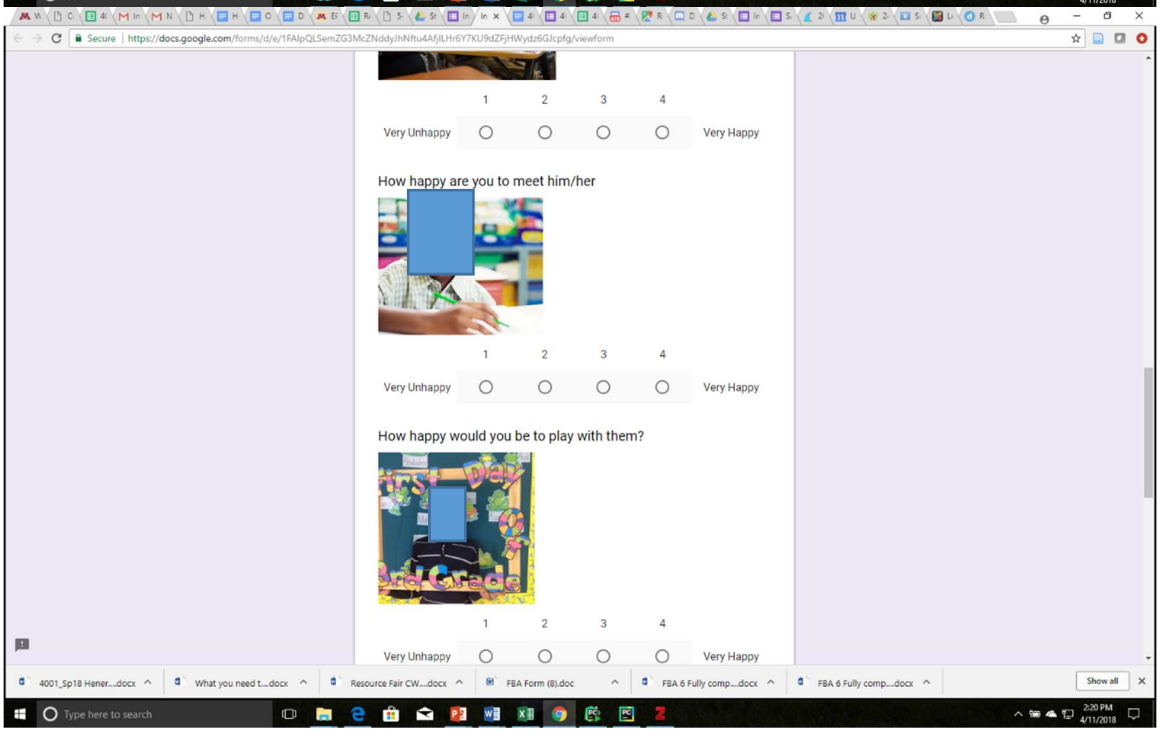
Google Forms

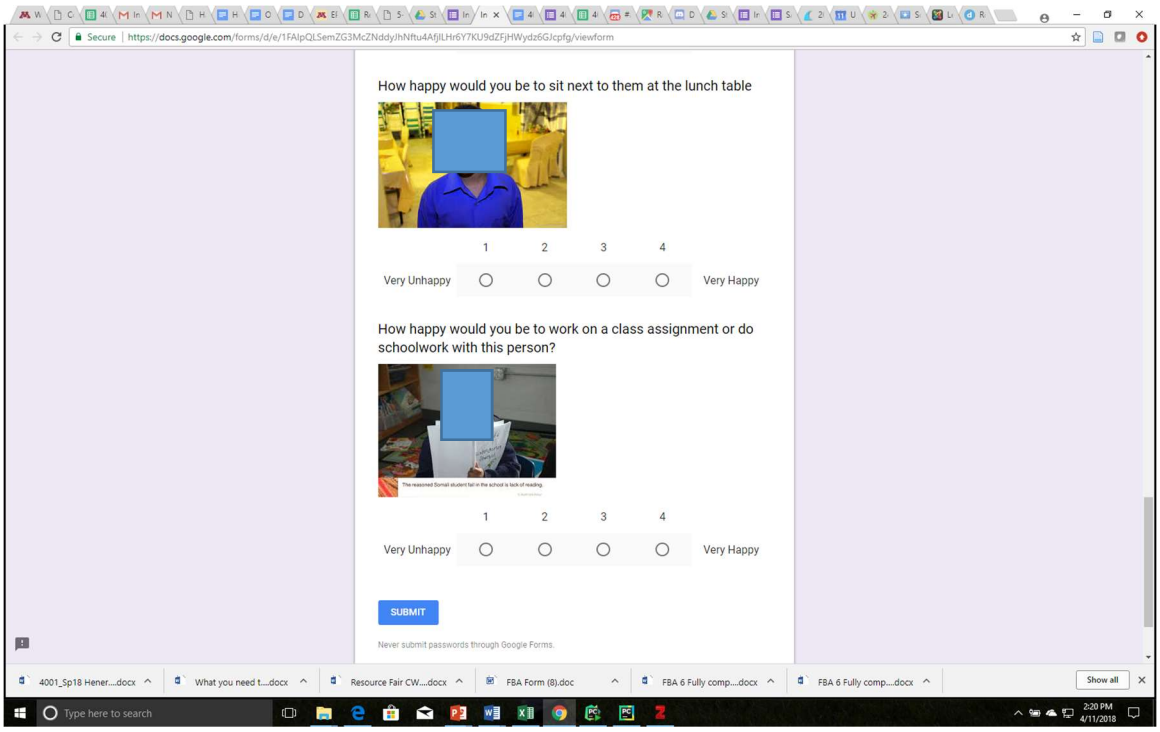
Appendix L

**INGROUP AND OUTGROUP BEHAVIOR INTENTIONS MEASURE (MALE) AS SEEN BY PARTICIPANTS**



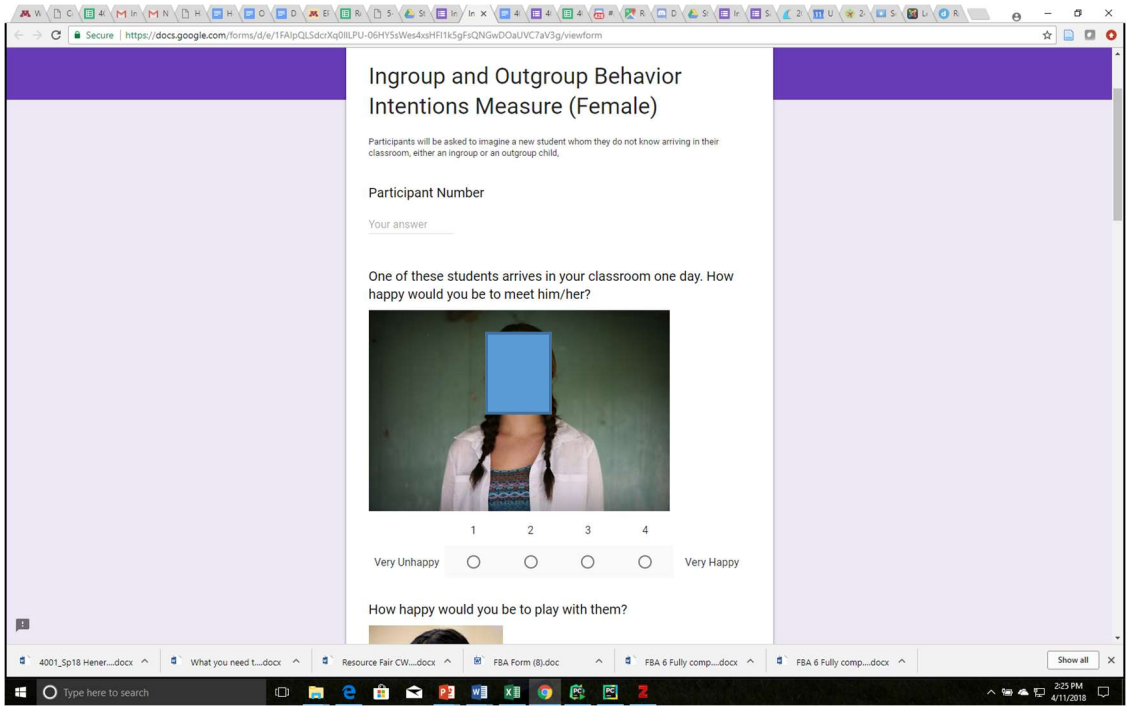

Secure
https://docs.google.com/forms/d/e/1FAIpQLSemZG3McZNddyJhNftu4AfjlLHr6Y7KU9dZFjHWydz6GJcpfg/viewform
1 2 3 4
Very Unhappy
Very Happy
How happy would you be to sit next to them at the lunch table
1 2 3 4
Very Unhappy
Very Happy
How happy would you be to work on a class assignment or do schoolwork with this person?
1 2 3 4
Very Unhappy
Very Happy
How happy are you to meet him/her
4001_Sp18 Hener....docx
What you need t....docx
Resource Fair CW....docx
FBA Form (8).doc
FBA 6 Fully comp....docx
FBA 6 Fully comp....docx
Show all
Type here to search
2:20 PM
4/11/2018

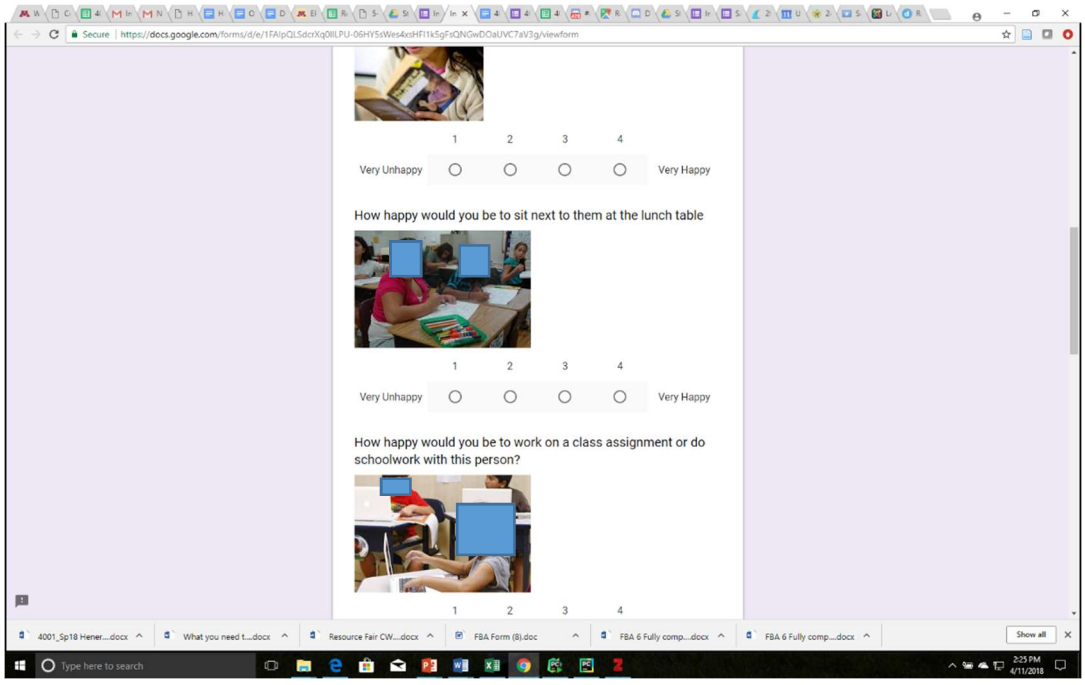

Secure
https://docs.google.com/forms/d/e/1FAIpQLSemZG3McZNddyJhNftu4AfjlLHr6Y7KU9dZFjHWydz6GJcpfg/viewform
1 2 3 4
Very Unhappy
Very Happy
How happy are you to meet him/her
1 2 3 4
Very Unhappy
Very Happy
How happy would you be to play with them?
1 2 3 4
Very Unhappy
Very Happy
4001_Sp18 Hener....docx
What you need t....docx
Resource Fair CW....docx
FBA Form (8).doc
FBA 6 Fully comp....docx
FBA 6 Fully comp....docx
Show all
Type here to search
2:20 PM
4/11/2018

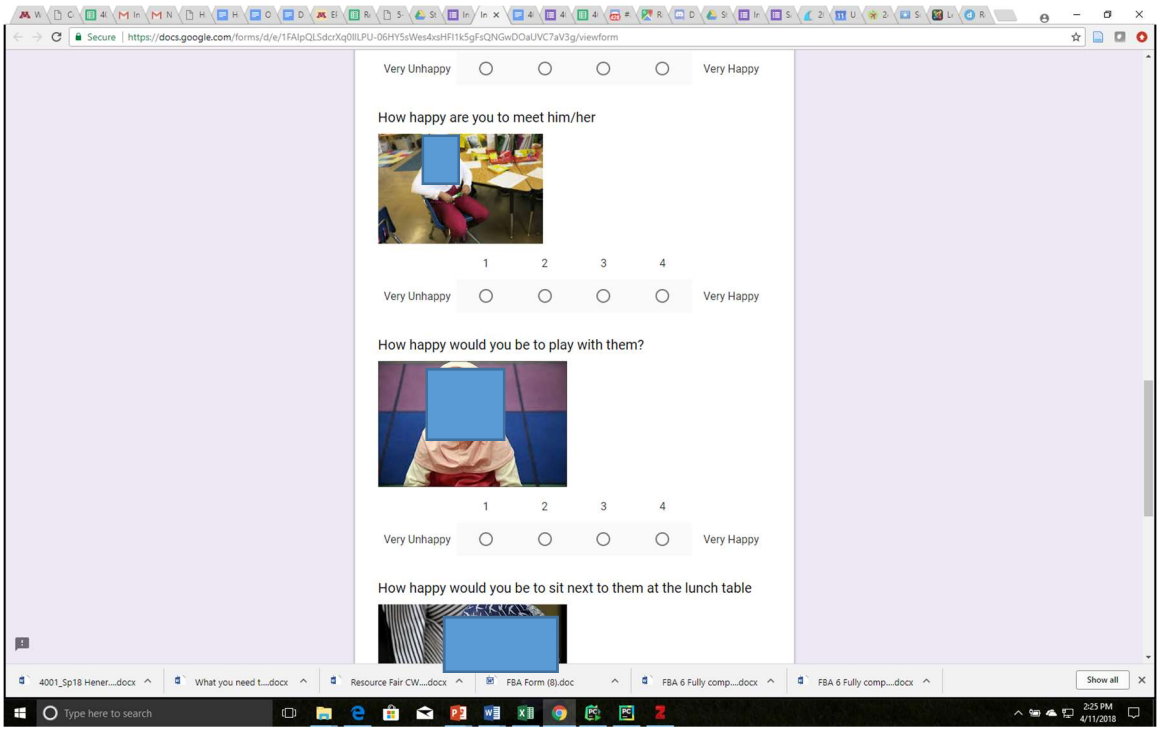

Secure https://docs.google.com/forms/d/e/1FAIpQLSemZG3McZNddyJhNftu4AfjlLHr6Y7KU9dZFjHWydz6GJcpfg/viewform
How happy would you be to sit next to them at the lunch table
1 2 3 4
Very Unhappy
Very Happy
How happy would you be to work on a class assignment or do schoolwork with this person?
1 2 3 4
Very Unhappy
Very Happy
SUBMIT
Never submit passwords through Google Forms.
4001_Sp18 Hener....docx
What you need t....docx
Resource Fair CW....docx
FBA Form (8).doc
FBA 6 Fully comp....docx
FBA 6 Fully comp....docx
Show all
Type here to search
2:20 PM
4/11/2018

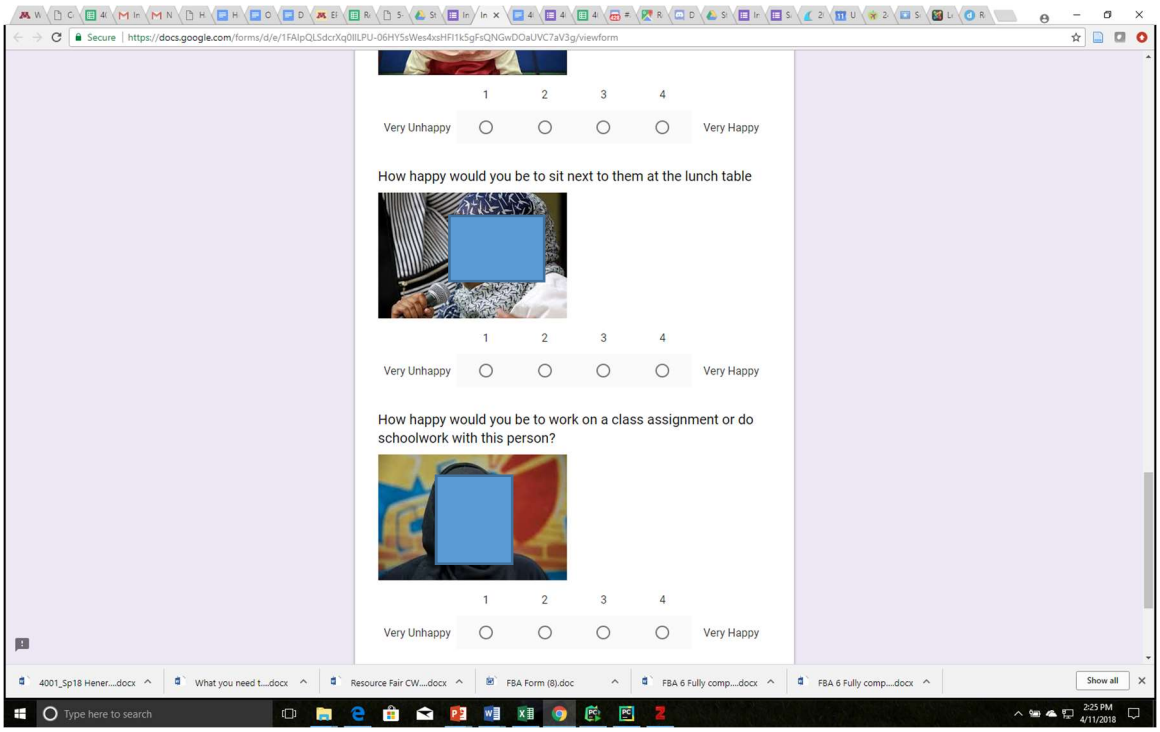
Appendix M

# INGROUP AND OUTGROUP BEHAVIOR INTENTIONS MEASURE (FEMALE) AS SEEN BY PARTICIPANTS



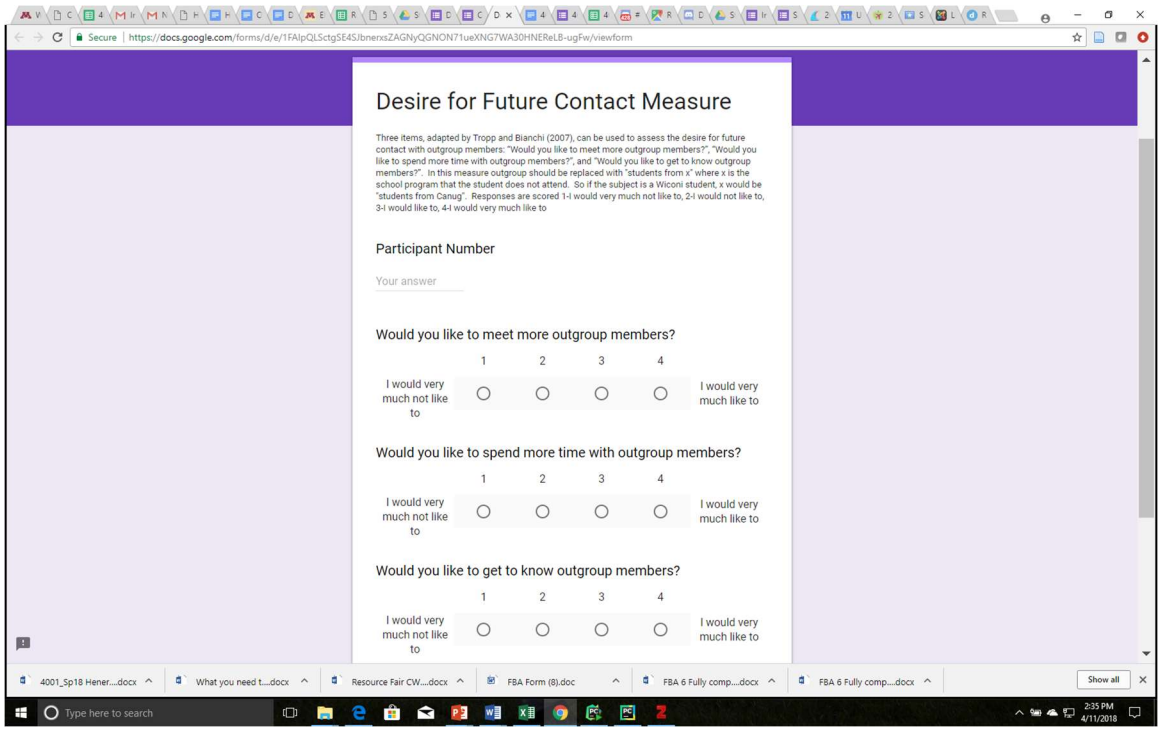




Secure https://docs.google.com/forms/d/e/1FAIpQLSdcrXq0IILPU-06HY5sWes4xsHFI1k5gFsQNGwDOaUVC7aV3g/viewform
Very Unhappy
Very Happy
How happy are you to meet him/her
1 2 3 4
Very Unhappy
Very Happy
How happy would you be to play with them?
1 2 3 4
Very Unhappy
Very Happy
How happy would you be to sit next to them at the lunch table
4001_Sp18 Hener....docx
What you need t....docx
Resource Fair CW....docx
FBA Form (8).doc
FBA 6 Fully comp....docx
FBA 6 Fully comp....docx
Show all
Type here to search
2:25 PM
4/11/2018


Secure https://docs.google.com/forms/d/e/1FAIpQLSdcrXq0IILPU-06HY5sWes4xsHFI1k5gFsQNGwDOaUVC7aV3g/viewform
1 2 3 4
Very Unhappy
Very Happy
How happy would you be to sit next to them at the lunch table
1 2 3 4
Very Unhappy
Very Happy
How happy would you be to work on a class assignment or do schoolwork with this person?
1 2 3 4
Very Unhappy
Very Happy
4001_Sp18 Hener....docx
What you need t....docx
Resource Fair CW....docx
FBA Form (8).doc
FBA 6 Fully comp....docx
FBA 6 Fully comp....docx
Show all
Type here to search
2:25 PM
4/11/2018

Appendix N

**DESIRE FOR FUTURE CONTACT MEASURE**



Desire for Future Contact Measure

Three items, adapted by Tropp and Bianchi (2007), can be used to assess the desire for future contact with outgroup members: "Would you like to meet more outgroup members?", "Would you like to spend more time with outgroup members?", and "Would you like to get to know outgroup members?". In this measure outgroup should be replaced with "students from x" where x is the school program that the student does not attend. So if the subject is a Wiconi student, x would be "students from Canug". Responses are scored 1-I would very much not like to, 2-I would not like to, 3-I would like to, 4-I would very much like to

Participant Number

Your answer

Would you like to meet more outgroup members?

| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| I would very much not like to | ○ | ○ | ○ | ○ | I would very much like to |

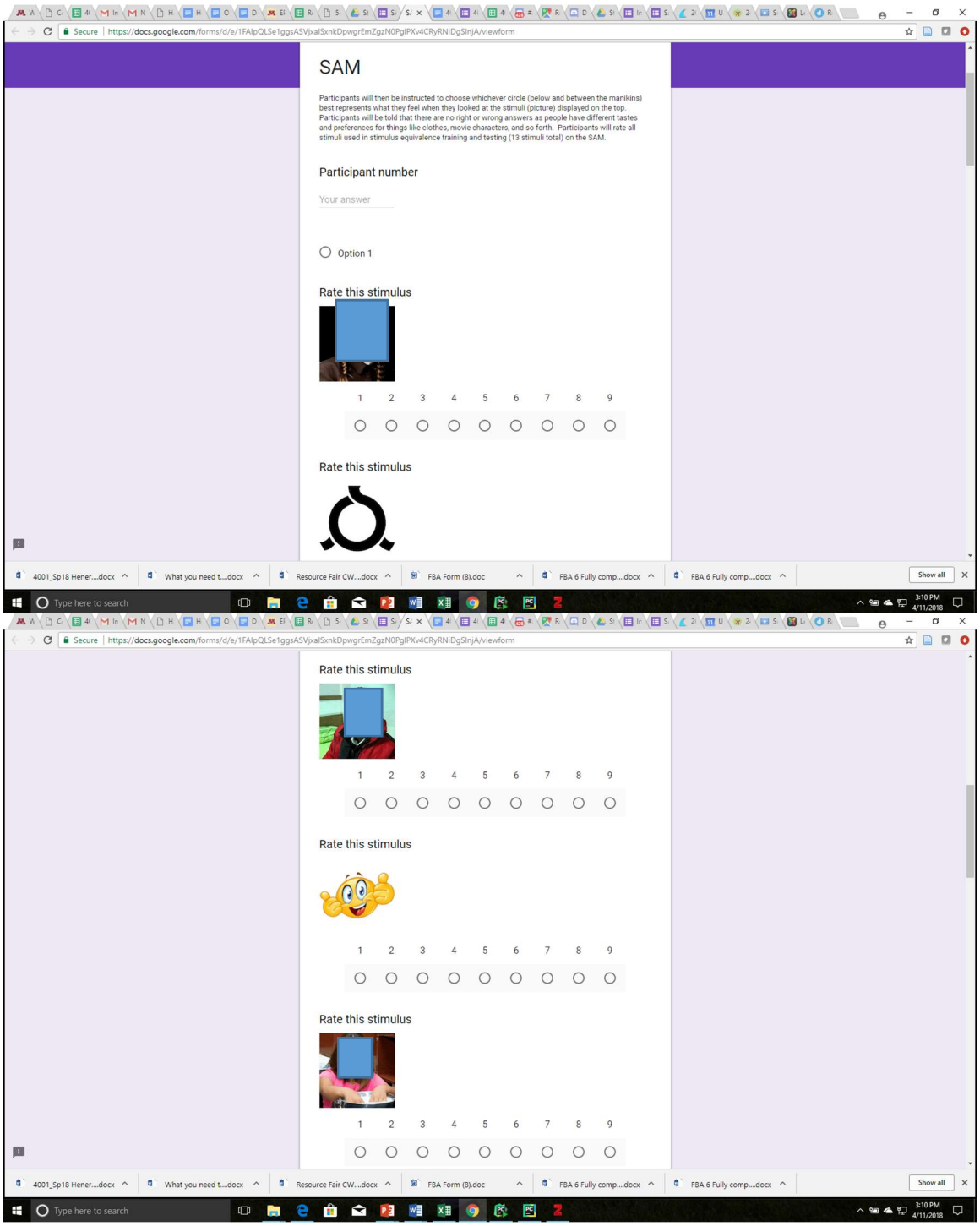
Would you like to spend more time with outgroup members?

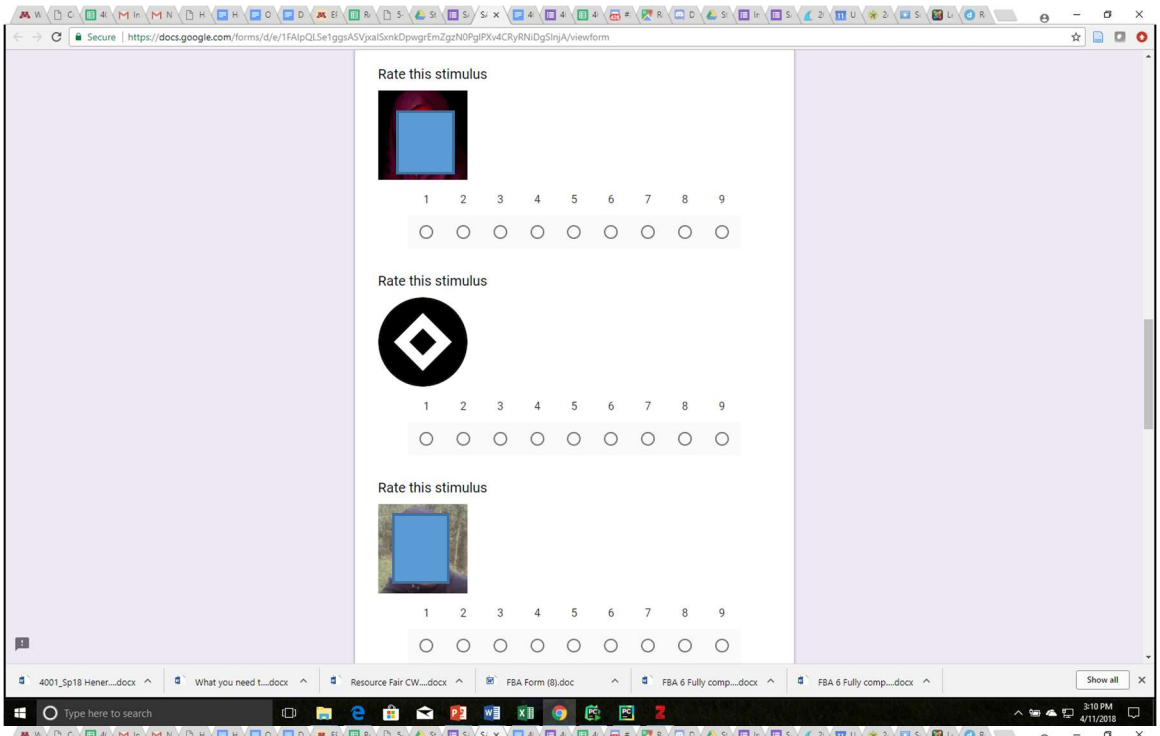
| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| I would very much not like to | ○ | ○ | ○ | ○ | I would very much like to |

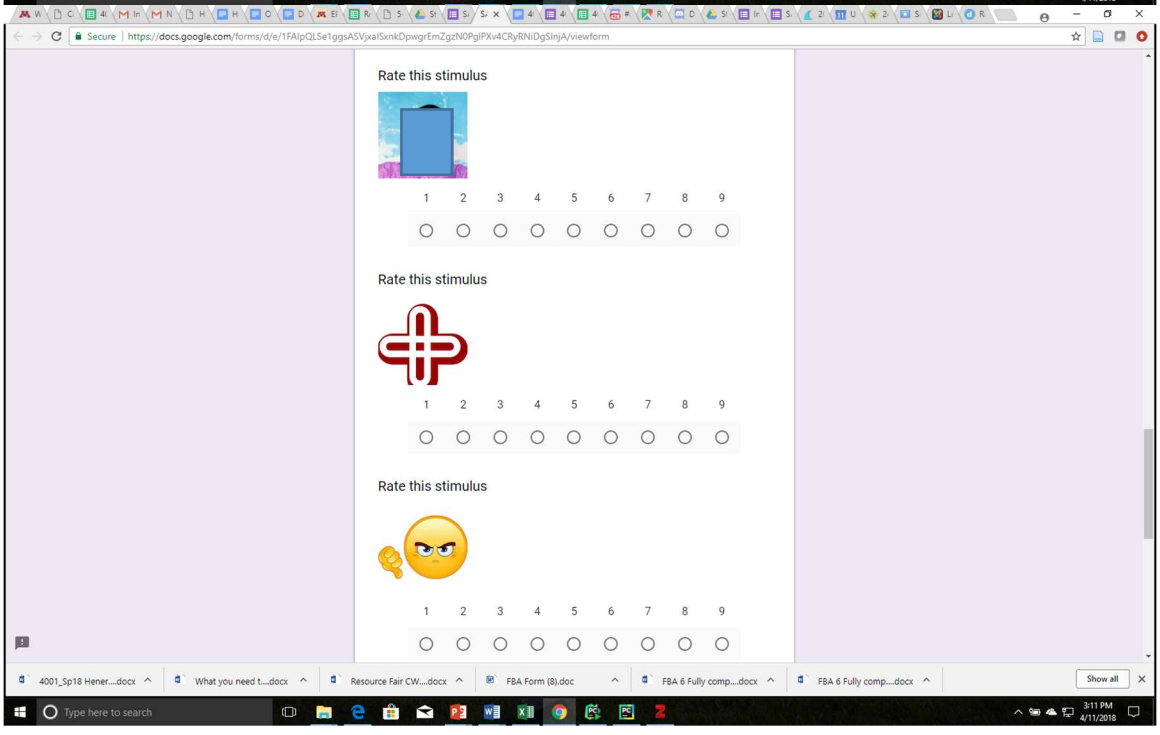
Would you like to get to know outgroup members?

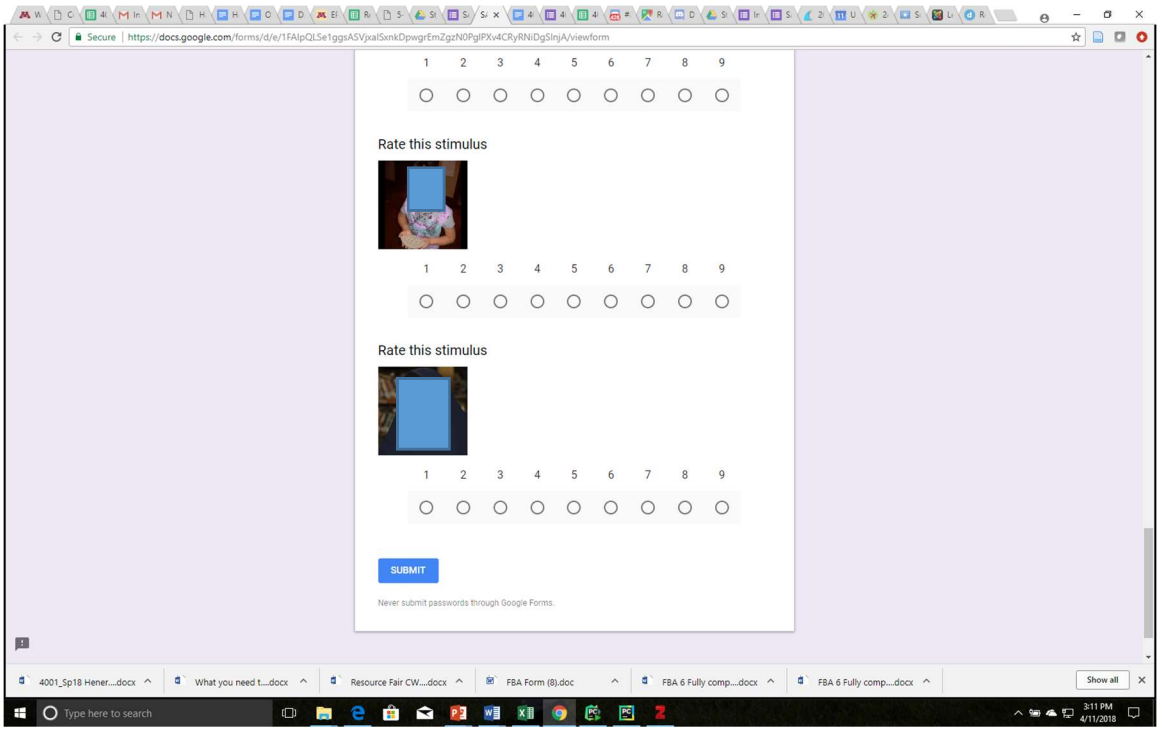
| | 1 | 2 | 3 | 4 | |
|---|---|---|---|---|---|
| I would very much not like to | ○ | ○ | ○ | ○ | I would very much like to |

# Appendix O
# SAM

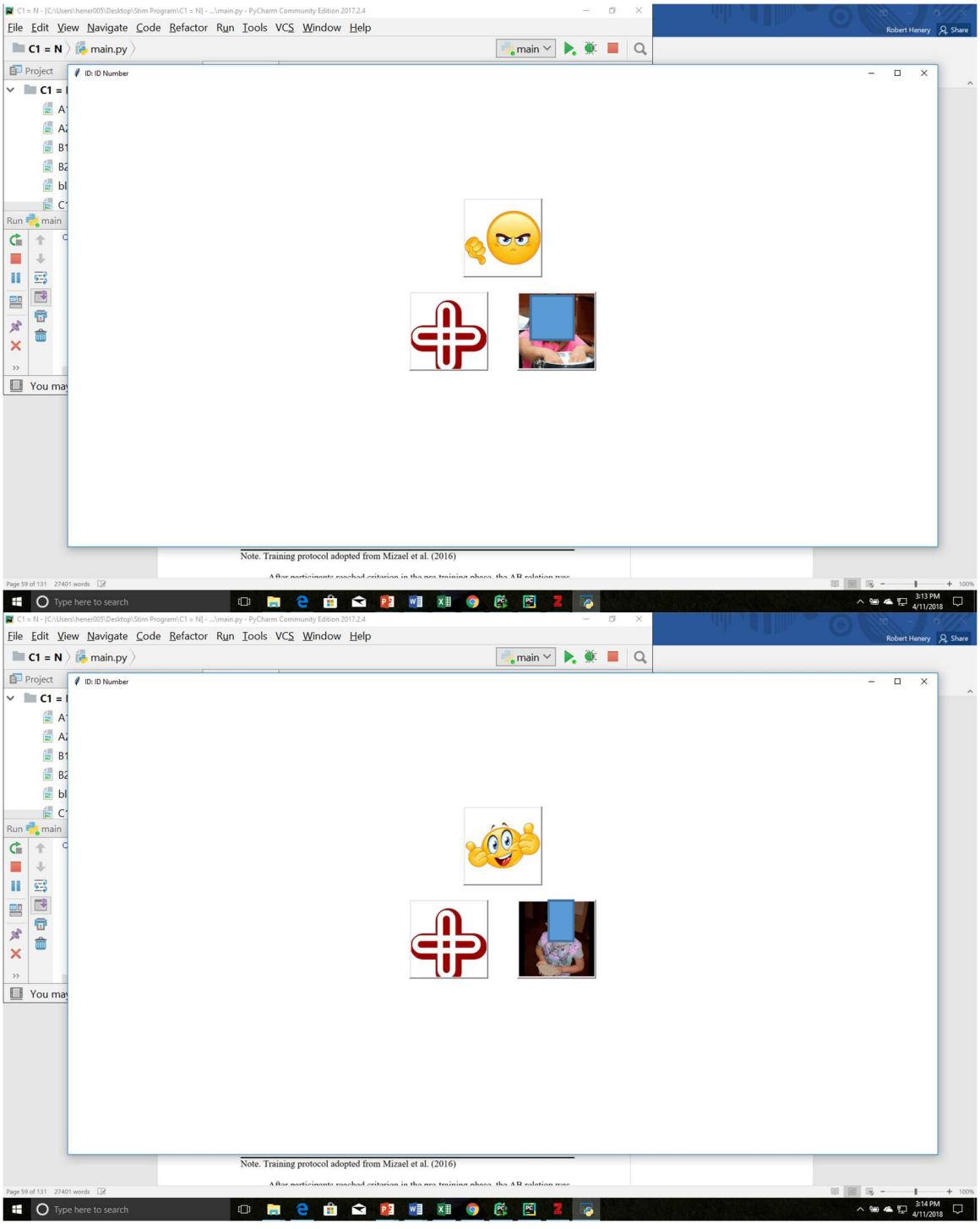



Secure | https://docs.google.com/forms/d/e/1FAIpQLSe1ggsASVjxaISxnkDpwgrEmZgzN0PgIPXv4CRyRNiDgSInjA/viewform
Rate this stimulus
1 2 3 4 5 6 7 8 9
Rate this stimulus
1 2 3 4 5 6 7 8 9
Rate this stimulus
1 2 3 4 5 6 7 8 9
4001_Sp18 Hener....docx
What you need t....docx
Resource Fair CW....docx
FBA Form (8).doc
FBA 6 Fully comp....docx
FBA 6 Fully comp....docx
Show all
Type here to search
3:10 PM
4/11/2018

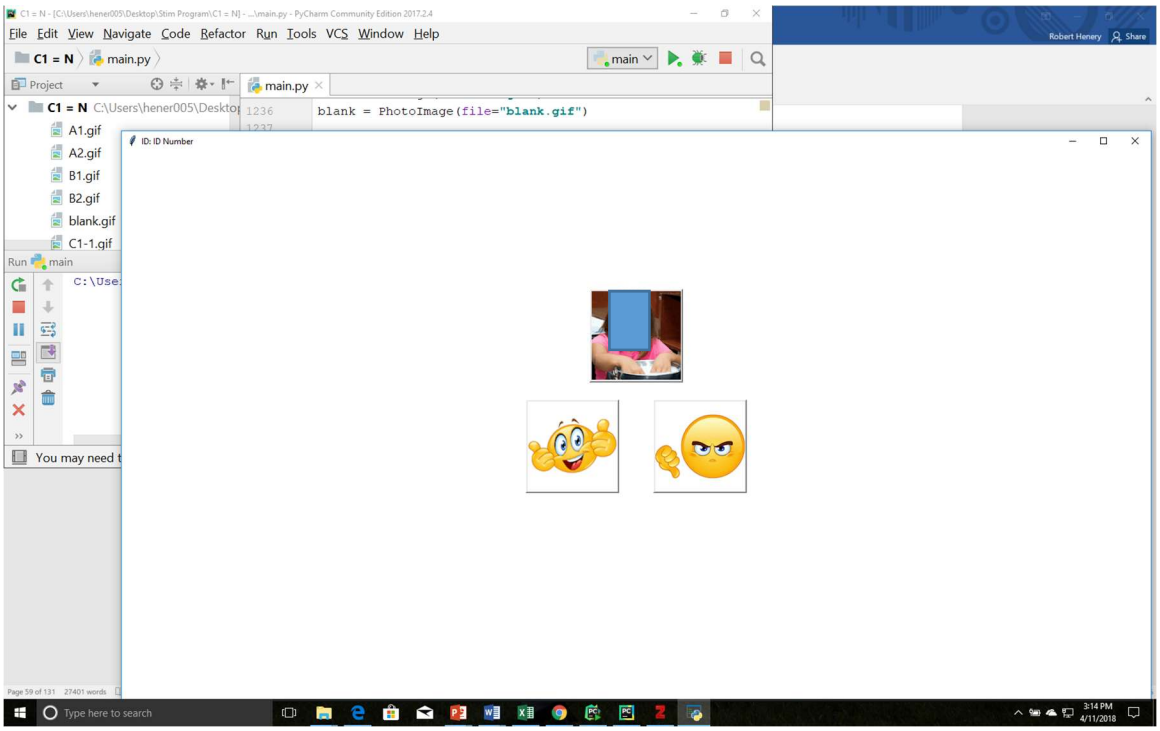

Secure | https://docs.google.com/forms/d/e/1FAIpQLSe1ggsASVjxaISxnkDpwgrEmZgzN0PgIPXv4CRyRNiDgSInjA/viewform
Rate this stimulus
1 2 3 4 5 6 7 8 9
Rate this stimulus
1 2 3 4 5 6 7 8 9
Rate this stimulus
1 2 3 4 5 6 7 8 9
4001_Sp18 Hener....docx
What you need t....docx
Resource Fair CW....docx
FBA Form (8).doc
FBA 6 Fully comp....docx
FBA 6 Fully comp....docx
Show all
Type here to search
3:11 PM
4/11/2018

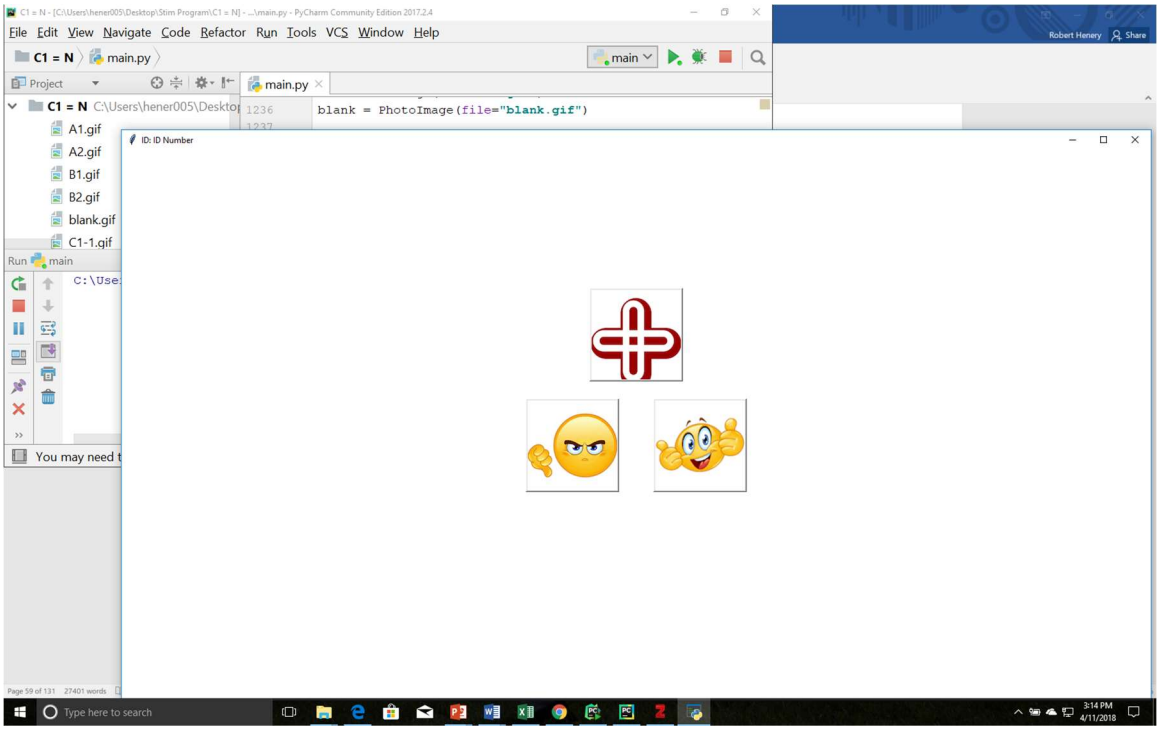

Secure https://docs.google.com/forms/d/e/1FAIpQLSe1ggsASVjxalSxnkDpwgrEmZgzN0PgIPXv4CRyRNiDgSInjA/viewform
1 2 3 4 5 6 7 8 9
Rate this stimulus
1 2 3 4 5 6 7 8 9
Rate this stimulus
1 2 3 4 5 6 7 8 9
SUBMIT
Never submit passwords through Google Forms.
4001_Sp18 Hener....docx
What you need t....docx
Resource Fair CW....docx
FBA Form (8).doc
FBA 6 Fully comp....docx
FBA 6 Fully comp....docx
Show all
Type here to search
3:11 PM
4/11/2018

# Appendix P
# AC TEST



# Appendix Q
# CA TEST



# Appendix R
# AC3 TEST